Inaugural dissertation
for
obtaining the doctoral degree
of the
Combined Faculty of Mathematics, Engineering and Natural Sciences
of the
Ruprecht - Karls - University
Heidelberg

Presented by
Zahra Abadi, M.Sc.
born in: Bandaranzali, Iran
Oral examination: 28th of March 2022

Loss of the Tumor Suppressive Function of PRSS23 Drives Hepatocellular Carcinoma Development

Referees:
Prof. Dr. rer. nat. Peter Angel
Prof. Dr. med. Thomas Longerich

To my family

Shohreh Reza Mahta Torben

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3D	Three-dimensional space
AFP	Alpha fetoprotein
AKT1	AKT serine/threonine kinase 1
ALD	Alcoholic-liver disease
AMPK	$5'$ AMP-activated protein kinase
ANAPC1	Anaphase-promoting complex subunit 1
ANNOVAR	ANNOtate VARiation
ANOVA	Analysis of variance
AP1B1	AP-1 complex subunit beta-1
AP2M1	AP-2 complex subunit mu
APC	Adenomatous polyposis coli protein
ARID1A	AT-rich interaction domain 1A
ARID2	AT-rich interaction domain 2
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
AXIN1	Axin 1
AXIN2	Axin 2
BAM	Binary Alignment Map
BCLC	Barcelona-Clinic-Liver Cancer
BSA	Bovine serum albumin
BWA	Burrows-Wheeler Aligner
CCNA2	Cyclin A2
CCND1	Cyclin D1
CCND2	Cyclin D2
CCNE1	Cyclin E1
CDH1	Cadherin 1
CDKN2A	Cyclin dependent kinase inhibitor 2A
CDKN2B	Cyclin dependent kinase inhibitor 2B
cDNA	Complementary DNA
Chek2	Checkpoint kinase 2

CL	Cirrhotic liver
CLMP	CXADR like membrane protein
CLPP	ATP-dependent Clp protease proteolytic subunit, mitochondrial
CLTA	Clathrin light chain A
CO_2	Carbon dioxide
COX-2	Cyclooxygenase 2
CREBBP	CREB binding protein
CRP	C-reactive protein
CTNNB1	Catenin beta 1
Ctrl	Control
dbSNP132	Single nucleotide polymorphism database build 132
DDA	Data Dependent Acquisition
DIP2C	Disco interacting protein 2 homolog C
DLK1	Delta like non-canonical notch ligand 1
DM	Diabetes mellitus
DMEM	Dulbecco's Modified Eagle Medium
DMOG	Dimethyloxallyl glycine
DMSO	Dimethyl sulfoxide
DN	Dysplastic Nodule
DNA	Deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	Dithiothreitol
E2F1	E2F transcription factor 1
ECAR	Extracellular acidification rate
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EIF2	Eukaryotic Initiation Factor 2
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
EMILIN-1	Elastin microfibril interfacer 1
EMT	Epithelial-to-mesenchymal transition

ABBREVIATIONS

EndoMT	Endothelial-to-mesenchymal transition
EPCAM	Epithelial cell adhesion molecule
EPN4	Clathrin interactor 1
EPS5400	Exome Sequencing Project 5400
ERK1/2	Extracellular signal-regulated kinase 1/2
ERα	Estrogen receptor α
ETC	Electron transfer chain
EtOH	Ethanol
FAAP100	FA core complex associated protein 100
FBN2	Fibrillin 2
FBS	Fetal bovine serum
FDR	False discovery rate
FFPE	Formalin-fixed paraffin-embedded
FGF19	Fibroblast growth factor 19
FOXO	Forkhead box O
FZD4	Frizzled class receptor 4
G0/G1 phase	Gap 0/Gap 1 phase
G2/M phase	Gap 2/Mitosis phase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GATK	Genome analysis toolkit
GB	Gallbladder
GEMM	Genetically engineered mouse model
GFP	Green fluorescent protein
glycoATP	glycolytic ATP
GMPS	GMP synthase [glutamine-hydrolyzing]
GSEA	Gene Set Enrichment Analysis
H&E	Hematoxylin and eosin
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCl	Hydrochloric acid
HCV	Hepatitis C virus

HDAC	Histone deacetylase
HDAC1	Histone deacetylase 1
HEK 293T	Human embryonic kidney 293T
HGDN	High-grade Dysplastic Nodule
HGF	Hepatocyte growth factor
HIF-1	Hypoxia-inducible factor-1
HIP1R	Huntingtin-interacting protein 1-related protein
HOGA1	4-hydroxy-2-oxoglutarate aldolase 1
HTRA1/2	High-temperature requirement A serine peptidase 1/2
HTVI	Hydrodynamic tail vein injection
iBAQ	intensity-based absolute quantification
IGF1R	Insulin like growth factor 1 receptor
IgG	Immunoglobulin G
IL6	Interleukin 6
InDel	Insertion-deletion
IP	Immunoprecipitation
IR	Inverted Repeat
IRES	Internal ribosome entry site
IRF2	Interferon regulatory factor 2
JAK	Tyrosine-protein kinase JAK
K19	Keratin 19
KCNQ2	Potassium voltage-gated channel subfamily q member 2
KEGG	Kyoto Encyclopedia of Genes and Genomes
KMT2D	Lysine methyltransferase 2D
KRAS	KRAS proto-oncogene, GTPase
LAMA1	Laminin subunit alpha 1
LB	Lysogeny broth
LGDN	Low-grade Dysplastic Nodule
logFC	log fold-change
LOH	Loss of heterozygosity
LONP1	Lon protease homolog, mitochondrial

MAP3K4	Mitogen-activated protein kinase kinase kinase 4
Mb	Megabase
MCM3	DNA replication licensing factor MCM3
MCM6	DNA replication licensing factor MCM6
MDM4	MDM4 regulator of p53
MEM NEAA	Minimum Essential Medium Non-Essential Amino Acids
MeOH	Methanol
MET	MET proto-oncogene, receptor tyrosine kinase
miR30	miR-30 microRNA
MiT/TFE	Microphthalmia/transcription factor E
mitoATP	mitochondrial ATP
MLL4	Lysine methyltransferase 2B (KMT2B)
mRNA	messenger ribonucleic acid
MS	Mass spectrometry
MSCV	Murine stem cell virus
MSMS	Tandem mass spectrometry
mTOR	Mammalian target of rapamycin
MTT	Methylthiazolyldiphenyl-tetrazolium bromide
MUT	Mutant
MYC	Myc proto-oncogene protein
NaCl	Sodium chloride
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
NCE	Normalized Collision Energy
NES	Normalized enrichment score
NFE2L2	NFE2 like bZIP transcription factor 2
NGS	Next generation sequencing
NMNAT2	Nicotinamide nucleotide adenylyltransferase 2
NP40	Nonidet P40
OCR	Oxygen consumption rate
OD	Optical density

OGT	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit
pAMPK	phosphorylated AMPK
PARP	Poly (ADP-ribose) polymerase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
PER	Proton efflux rate
pERK1/2	phosphorylated ERK1/2
PFA	Paraformaldehyde
pH	Potential of hydrogen
PI	Propidium iodide
PI3K	Phosphoinositide 3-kinases
PICALM	Phosphatidylinositol binding clathrin assembly protein
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PIK3CG	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma
PLD1	Phospholipase D1
PMSF	Phenylmethylsulfonyl fluoride
PolyPhen-2	Polymorphism Phenotyping v2
pPKG	Mouse phosphoglycerate kinase 1 promoter
PRPH	Peripherin
PRSS23	Serine protease 23
RB1	Retinoblastoma 1
RE	Restriction enzyme
RNA	Ribonucleic acid
RNAi	RNA interference
Rot/AA	Rotenone plus antimycin A
RPS6KA3	Ribosomal protein s6 kinase A3
RT	Room temperature
RTN4IP1	Reticulon 4 interacting protein 1

ABBREVIATIONS

S phase	Synthesis phase
S.O.C	Super Optimal Catabolite
SB13	Sleeping beauty 13
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEM	Standard error of the mean
shRNA	Short hairpin RNA
SNP	Single nucleotide polymorphism
STAT	Signal transducer and activator of transcription
STON2	Stonin 2
SV40	Simian vacuolating virus 40
TBST	Tris-buffered saline with Tween 20
TCGA	The Cancer Genome Atlas
TE-Buffer	Tris-acetate-EDTA buffer
TERT	Telomerase reverse transcriptase
TFA	Trifluoroacetic acid
TGF-α	Transforming growth factor α
TGF-β	Transforming growth factor beta
TMPRSS15	Transmembrane protease serine 15
TMPRSS2/4	Transmembrane protease serine 2/4
TP53	Tumor protein 53
Trp53	Transformation related protein 53
TSC1/2	TSC complex subunit 1/2
UT	Untreated
V	Voltage
VEGF	Vascular endothelial growth factor
VEGFA	Vascular endothelial growth factor A
VEGFR2	Vascular endothelial growth factor receptor 2
VF	Variant frequency
VSN	Variance Stabilization Normalization
VWA5A	von Willebrand factor A domain containing 5A

WB	Western blot
WES	Whole exome sequencing
WT	Wild-type

ABBREVIATIONS

SUMMARY

Hepatocellular carcinoma (HCC) is the most prevalent type of primary liver cancer and belongs to the four most lethal malignancies worldwide. HCC constitutes a major health concern as it is frequently diagnosed at late-stage of the disease and the presence of concomitant cirrhosis makes curative surgery inapplicable for most of the patients and concurrently reduces the options for systemic therapy. In fact, therapy is most effective during early stages of the disease. Hence, early diagnosis and treatment are of prime importance.

Human hepatocarcinogenesis represents a step-wise process, in which accumulation of somatic mutations allows for clonal selection of driver gene mutations, which finally results in malignant transformation of premalignant Dysplastic Nodule into HCC. In the presented study, we aimed: (i) to identify and validate driver mutations of hepatocarcinogenesis; and (ii) to investigate the molecular mechanism through which they promote malignant transformation.

The present thesis describes the identification of clonally expanded driver mutations of hepatocarcinogenesis via integration of whole exome sequencing and *in vivo* RNAi screening. *In vivo* shRNA screening of the identified clonally expanded variants using a mosaic mouse model of HCC revealed the tumor suppressive role of PRSS23 in the context of activated MYC and AKT1 signaling. Subsequent *in vivo* validation experiments using the individual variants confirmed the oncogenic potential of the identified PRSS23 variant (P230A) in MYC overexpressing *Trp53* heterozygous mice. 3D computational modelling of PRSS23 located the amino acid substitution P230A in close proximity to the catalytic site of the protein suggesting a potential interfering effect on the enzymatic activity.

Using a primary murine HCC cell line derived from *Trp53* heterozygous mouse expressing MYC and activated AKT1 together with the wildtype or P230A variant of PRSS23, I demonstrated that overexpression of PRSS23^{P230A} elevates the proliferation rate and cell viability of HCC cells compared to PRSS23^{WT} overexpressing control cells, and promotes the invasiveness of these cells. The results of this thesis propose a previously unrecognized role for PRSS23 in mitochondria of HCC cells. Real-time analysis of the ATP production rate and morphological analysis of mitochondria in HCC cells demonstrated that overexpression of PRSS23^{P230A} contributes to a metabolic switch towards enhanced oxidative phosphorylation likely via a process involving remodeling of mitochondrial cristae. Remarkably, transcriptomic analysis of HCC cells expressing wildtype or the P230A variant of PRSS23 unveiled differential regulation of several signaling pathways. In particular, MYC activation was

elevated in PRSS23^{P230A} expressing cells. In addition, enrichment of PRSS23 interacting proteins in these cellular processes further corroborated the latter finding. Strikingly, evaluating the activation level of MYC in HCC cells validated that PRSS23^{P230A} augments MYC activation.

All in all, the results of this thesis demonstrate that the P230A mutation of PRSS23 promotes malignant transformation during hepatocarcinogenesis by promoting MYC activity in HCC cells involving structural and functional remodeling of mitochondria.

ZUSAMMENFASSUNG

Das Hepatozelluläre Karzinom (HCC) ist die häufigste Form des primären Leberkrebses und gehört zu den vier tödlichsten Krebsarten weltweit. Es stellt ein großes Gesundheitsproblem dar. Einerseits wird es häufig erst spät diagnostiziert. Andererseits schließt das gleichzeitige Vorliegen einer Leberzirrhose bei der Mehrzahl der Patienten einen kurativen chirurgischen Eingriff aus und schränkt zudem die Möglichkeiten der systemischen Therapie ein. Am effektivsten kann das HCC in frühen Erkrankungsstadien therapiert werden, weshalb eine frühe Diagnose und Behandlung wesentlich sind.

Die humane Hepatokarzinogenese stellt einen mehrstufigen Prozess dar, bei dem die Akkumulation somatischer Mutationen die klonale Selektion von Treibergenmutationen ermöglicht, welche schließlich die maligne Transformation eines prämalignen dysplastischen Knoten (DN) in ein HCC bewirken. Die vorliegenden Arbeit hat zum Ziel: (i) die Treibermutationen der Hepatokarzinogenese zu identifizieren und zu validieren; und (ii) den molekularen Mechanismus zu untersuchen, durch den die maligne Transformation begünstigt wird.

Es wurden die klonal expandierten Treibermutationen der Hepatokarzinogenese mittels Exomsequenzierung identifiziert und durch *in vivo* RNAi-Untersuchungen validiert. Hierzu wurde ein Mosaik-Mausmodell des HCC verwendet. Mit dessen Hilfe konnte eine tumorsuppressive Rolle von PRSS23 im Kontext eines aktiven MYC- und AKT1- Signalwegs nachgewiesen werden. Nachfolgende *in vivo* Experimente, bei denen die einzelnen Genvarianten exprimiert wurden, bestätigten das tumorfördernde Potenzial der identifizierten Variante PRSS23^{P230A} in MYC überexprimierten und *Trp53*-heterozygoten Maushepatozyten. Ein 3D-Computermodell von PRSS23 lokalisierte die Position der varianten Aminosäure P230A in unmittelbarer Nähe zum katalytischen Zentrum, wodurch die enzymatische Aktivität des Proteins beeinflusst werden könnte.

Unter zur Hilfenahme einer primären murinen HCC-Zelllinie, die aus *Trp53*-heterozygoten Mäusen unter Verwendung eines *Myc*- und *MyrAKT1*-exprimierenden Transposonvektors generiert wurde, konnte ich zeigen, dass durch die Überexpression von PRSS23^{P230A} die Proliferationsrate und Überlebensfähigkeit von HCC-Zellen im Vergleich zu PRSS23^{WT} erhöht ist und zudem die Invasionsfähigkeit der Tumorzellen gesteigert wird. Die Ergebnisse dieser Arbeit zeigen zudem eine bis dato unbekannte Funktion von PRSS23 in den Mitochondrien von HCC-Zellen auf. Echtzeitanalysen der ATP-Produktionsrate und die

morphologische Untersuchung der Mitochondrien in den HCC-Zellen zeigten, dass die Überexpression von PRSS23^{P230A} zu einem metabolischen Switch hin zu einer verstärkten oxidativen Phosphorylierung beiträgt. Dies beruht wahrscheinlich auf einem zellulären Prozess, welcher zum Umbau der mitochondrialen Cristae führt. Bemerkenswerterweise zeigte die Transkriptomanalyse der HCC-Zellen, die entweder den Wildtyp oder die P230A Variante von PRSS23 exprimierten, eine differentielle Regulation einiger Signalkaskaden. Insbesondere die Aktivität von MYC war in den PRSS23^{P230A} –exprimierenden Zellen erhöht. Die Anreicherung von Proteinen, die mit einem aktivierten MYC-Signalweg assoziiert sind, in Immunopräzipitaten von PRSS23^{P230A}-exprimierenden Zellen, passt zu diesem Gesamtbild. Die verstärkte Aktivierung des MYC-Signalwegs durch PRSS23^{P230A} wurde zudem zweifelsfrei durch einen Reporterassay belegt.

Insgesamt zeigen die Ergebnisse dieser Arbeit, dass die PRSS23^{P230A}-Mutation die maligne Transformation von Hepatozyten begünstigt, indem einerseits das Aktivitätsprofil von MYC potenziert und zum anderen durch einen funktionellen und strukturellen Umbau der Mitochondrien die Energieversorgung der Tumorzellen optimiert wird.

1 INTRODUCTION

1.1 Hepatocellular carcinoma

1.1.1 Etiology and risk factors

Hepatocellular carcinoma (HCC) is the most prevalent type of primary liver cancer accounting for approximately 90% of all cases. Chronic liver disease leading to cirrhosis is present in the majority of HCC patients. Hepatitis B virus (HBV) and hepatitis C virus (HCV) infection, chronic alcohol abuse and non-alcoholic steatohepatitis (NASH) are the main etiological factors preceding HCC development worldwide. In Asia and Africa, HBV infection is the prevailing risk factor accounting for 60% of HCC cases. [1] In Africa, higher probability of exposure to aflatoxin B1 escalates the risk of HCC development due to its carcinogenic effects acting synergistic with HBV infection. [2] In the Western world, HBV accounts for only 20% of HCC while chronic HCV infection constitutes a prevalent underlying etiology for emergence of HCC in North America, Europe, and Japan. [1] Excessive alcohol intake results in alcoholic-liver disease (ALD), which may induce liver cirrhosis followed by HCC development. It accounts for 30% of HCC cases worldwide. Similar to aflatoxin B1, chronic alcohol consumption acts synergistically with other risk factors of HCC including diabetes mellitus (DM) type 2 and viral hepatitis and may thus also increase the risk of HCC development in a comorbid constellation. [3]

Obesity and diabetes mellitus type 2 are closely associated with the evolution of non-alcoholic fatty liver disease (NAFLD). [4, 5] NAFLD comprises a spectrum of liver diseases ranging from simple steatosis to NASH with progressive liver fibrosis. NASH is the most rapidly growing etiological factor for HCC development. [6, 7] Despite the fact that cirrhotic liver is the typical microenvironment for HCC development, nearly 20% of HCC cases arise in non-cirrhotic livers from which almost half of them are attributed to NASH. [8, 9] All in all, it is well awaited that both alcoholic and non-alcoholic steatohepatitis will be the most relevant etiological factors driving the development of liver cirrhosis and HCC in the near future. [10]

In addition to these major risk factors for HCC development, gender and aging as well as ethnicity and demographic regions are also highlighted as factors conferring an increased risk of HCC development. HCC is much more prevalent in men than in women (ratio about 4:1). In addition, a large cohort study in the United States of America showed that women had less-advanced HCC and higher overall survival rates compared to men, which could be attributed to a higher frequency of NASH-related HCC development in non-cirrhotic livers in

men. [11] The incidence of HCC steadily increases in individuals older than age 60. [12] Disparities in HCC incidence is also seen regarding race and ethnicity, which may be related to differential incidence of predisposing single-nucleotide variants and/or the health care and insurance status of social minorities. [1, 13]

1.1.2 Epidemiology

With 830 000 deaths reported in 2020, liver cancer represents one of the four most common causes of cancer related death worldwide. [14] The overall 5-year survival rate of liver cancer patients remains poor. However, the incidence of liver cancer has recently stabilized compared to earlier observation periods, in which an increasing incidence was noted. The earlier increase in liver cancer incidence was mainly due to an increase in prevalence of NAFLD as well as the steadily aging population with higher prevalence of chronic HCV infection. Indeed, trends in liver cancer incidence portray preceding changes in the exposure to risk factors driving disease development. [15] The most prominent example is the striking fall in incidence of viral-induced HCC following universal HBV vaccination and the development of highly effective HCV treatment using direct-acting anti-viral therapies. [16, 17] Meanwhile, the incidence of NASH and alcohol-associated HCC is on rise not only in developed but also in developing countries with economic growth since high fat diet, lack of regular exercise, and excessive alcohol consumption become more and more prevalent behavioral patterns worldwide, thus, constituting an important future health problem. [18-20] The prevalence of NAFLD in the world´s population is estimated to be one fourth and the annual incidence of HCC among patients with NASH ranges from 0.5% to 2.6%. [6, 7]

1.1.3 Treatment

The clinical management of HCC patients continues to remain a major challenge despite all the recent advances in treatment options. In particular, the fact that about 80-90% of HCC cases develop in cirrhotic livers limits the spectrum of applicable therapeutic options. [21] Indeed, the overall performance status of the HCC patient and its liver function are the main parameters determining the treatment option of choice. [1] The 5-tired Barcelona-Clinic-Liver Cancer (BCLC) classification is mostly used to select the suitable treatment modality for a given patient (Figure 1). [22] Based on this staging system, patients diagnosed in early disease stages (BCLC 0 and A) are treated by either local tumor ablation, tumor resection, or liver transplantation. These potentially curative therapies notably double the median survival rates from 36 to 60 months and more compared to untreated patients. Patients with intermediate

2

stage tumors (BCLC B) and compensated hepatic function can benefit from trans-arterial chemoembolization, which has shown to improve the overall survival to 26-30 months. [22, 23] Advanced stage patients (BCLC C) are eligible for systematic treatment including atezolizumab plus bevacizumab [24, 25] and the tyrosine kinase inhibitors sorafenib [26, 27] and lenvatinib [28-30] as the first line, and regorafenib [31, 32], cabozantinib [33, 34], and the anti-VEGFR2 antibody ramucirumab [35] as the second- or third line therapies. Importantly, the combination of the checkpoint inhibitor atezolizumab and the VEGF inhibitor bevacizumab did not only improve the survival (67% in the combination group alive following one year of treatment compared to 55% in the sorafenib group) but also improved the quality of life of treated patients in the IMbrave150 trial. [25] Palliative care including nutritional and psychological support and mitigation of pain is the only remedy at hand for patients diagnosed with end stage disease (BCLC D). [22]
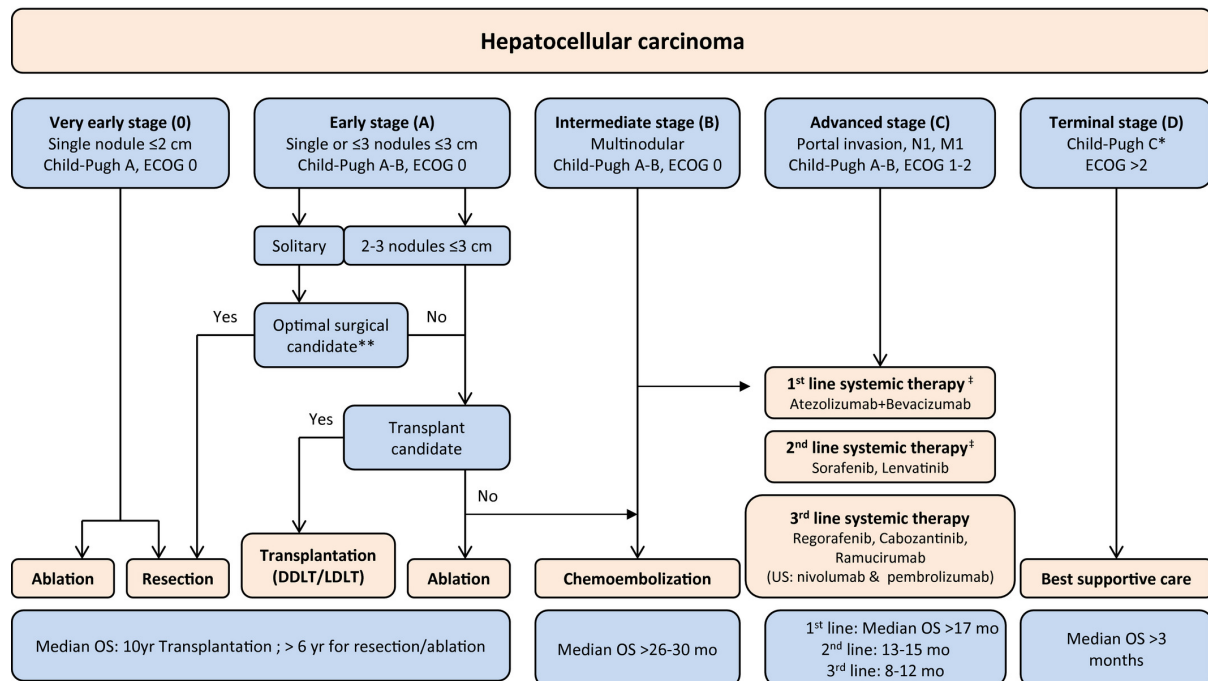
Figure 1. Descriptive schematic overview of the BCLC staging system and effective therapies for advanced stage disease. (modified and updated from EASL guidelines [36]). [23]

1.2 Step-wise process of hepatocarcinogenesis

Human hepatocarcinogenesis comprises a multi-step process lasting several decades. The cellular origin of hepatocellular carcinoma is still a matter of some debate. Nevertheless, compelling evidence confirm mature hepatocytes and less frequently hepatic progenitor cells (located in the canals of Hering) as the cellular origins of HCC. [37] As outlined above, liver

cirrhosis usually precedes HCC development. [21] In this precancerous microenvironment, premalignant Dysplastic Foci (< 1 mm in diameter) may evolve and may grow to become Dysplastic Nodules (DN, ≥ 1 mm in diameter). [38, 39] DN can be subclassified into low-grade (LGDN) and high-grade Dysplastic Nodules (HGDN). [39] HGDN carry a four-times increased risk of malignant transformation into early HCC and subsequent progression into advanced HCC. [40] A peculiar nodule-in-nodule appearance is occasionally found and is considered a proof of malignant transformation. In this scenario, either a HGDN may contain a clearly demarcated subnodule of early well-differentiated HCC or a well-differentiated HCC may contain a separate less differentiated tumor nodule. In the former case, the nodule-in-nodule appearance indicates malignant transformation while in the latter case it portrays tumor progression from early into advanced HCC. [41, 42] Thus, nodule-in-nodule growth is a morphological mirror of step-wise human hepatocarcinogenesis (Figure 2).
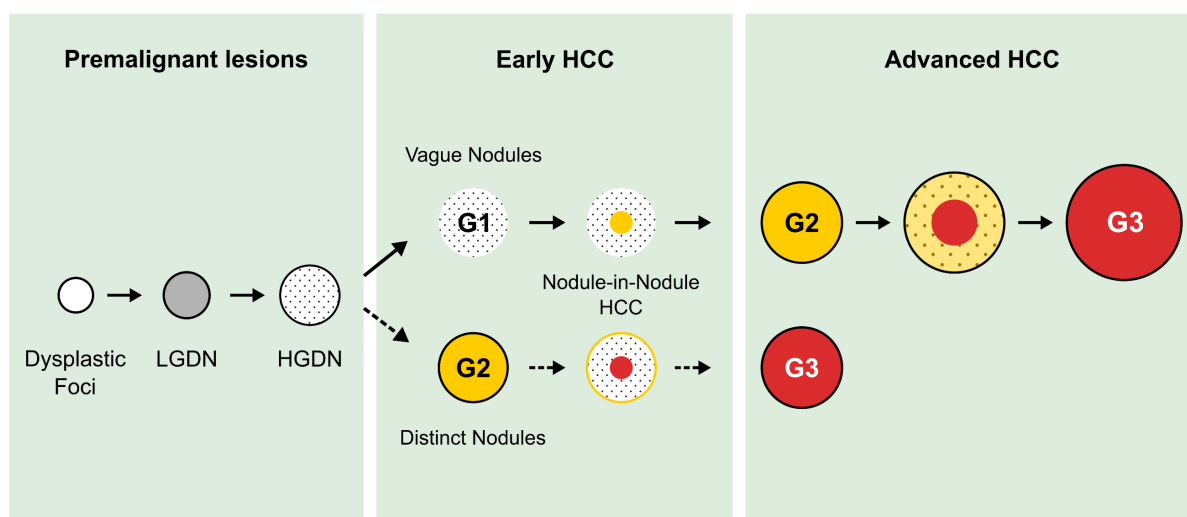
Figure 2. Model of multistep process of human hepatocarcinogenesis. (Modified from the International Consensus Group for Hepatocellular Neoplasia [41]). The scheme depicts the individual focal hepatocellular lesions observed during step-wise human hepatocarcinogenesis. [43] Abbreviations: LGDN, low-grade Dysplastic Nodule; HGDN, high-grade Dysplastic Nodule; G1, well/highly differentiated HCC; G2, moderately differentiated HCC; G3, poorly differentiated HCC.

1.3 Molecular alterations in HCC

Hepatocellular carcinoma comprises a genetically heterogenous disease which accumulates loads of passenger and driver gene mutations in the course of its development. Cancer driver mutations are genetic alterations which favor tumor growth from initiation to progression. These genomic alterations affect numerous pivotal cellular processes like proliferation, metabolism, stress response, migration and invasion as well as mechanisms of

immune escape; all of them constituting hallmarks of cancer and providing a growth and survival advantage to the tumor cells. On the contrary, passenger mutations develop in parallel but they are considered to neither promote nor impair the growth of tumor cells. [44] Over time, the mutational spectrum may differ between individual regions of an HCC nodule leading to a molecular heterogenous disease. In this evolutionary process, the cocktail of alterations providing the most optimal features in terms of growth, metabolism, migration, and invasion will survive and expand. As a consequence, so-called trunk alterations (occurring early during tumorigenesis) are shared by related or metastatic lesions. [45] Meanwhile, integrative genomic and transcriptomic meta-analysis of several independent human HCC cohorts has revealed that somatic alterations are not randomly distributed over the genome and that most individual alterations are rare, but indeed are enriched in certain molecular pathways allowing for molecular HCC classification (see below). [46]

1.3.1 Molecular classification

Over the past two decades, numerous molecular classifications of human HCC have been proposed. In a two-tiered system, hepatocellular carcinoma can be classified into a proliferative and a non-proliferative class each of which accounting for ~50% of HCC. [47-55] In contrast, the Boyault classification discriminates six subgroups and represents the most diversified system currently available. Others like the cancer genome atlas (TCGA) classifier discriminate three molecular HCC subgroups. [50, 51] Some of the most known molecular classifications and their association with clinical, etiological, histological, and (epi-)genetic features as well as altered signaling pathways and immunological features are summarized in Figure 3. [1] Proliferative class HCCs are characterized by *TP53* mutations and amplifications of *FGF19* and *CCND1*, as well as chromosomal instability and reduced global DNA methylation. Moreover, activation of cell cycle, PI3K-AKT-mTOR, RAS-MAPK, and MET signaling pathways are typically found in this class. [48-51, 54, 55] Tumors belonging to the proliferative class are in many cases poorly differentiated and therefore are more aggressive. This, translates into a higher probability of vascular invasion and worse outcome compared to other molecular subtypes. [55] The proliferative class can be further divided into two subclasses: the progenitor cell and the TGFβ-Wnt subclasses which more or less correspond to cluster 1 and cluster 3 of the TCGA classification. [51] The progenitor cell subclass represents 25-30% of HCC and is characterized by inactivating mutations of *AXIN1* and *RPS6KA3* and activation of AKT and IGF1R signaling pathways. These tumors are featured by expression of progenitor markers like K19, EPCAM, AFP, and activated MAPK pathway which may be

demonstrated by immunohistochemical detection of hyperphosphorylated ERK. [55] The TGFβ-Wnt subclass represents another 20% of HCC and is principally recognized via activation of non-canonical Wnt Signaling. [56] This subgroup can be further divided into G2 and G3 subgroups of the Boyault classifier. [50] Some of the HCCs belonging to G3 reveal a distinctive histomorphological phenotype called macrotrabecular-massive HCC. [57] Tumors belonging to this subgroup may show a high proliferation and mutations in the *TSC1* and *TSC2* genes or amplification of the *FGF19/CCND1* gene locus promoting a more aggressive phenotype. [1, 58]

Well differentiated HCCs are more likely to belong to the non-proliferative class (corresponding to cluster 2 of the TCGA classification). They are considered less aggressive and show less chromosomal aberrations compared to their proliferative class counterparts. [48-51, 54, 55] In the Boyault classification, the group of non-proliferative HCC is further divided into three subclasses (G4 to G6). The G4 HCCs may microscopically show a steatohepatitic phenotype, overexpression of C-reactive protein (CRP), and activation of the IL6-JAK-STAT signaling cascade. [50] Among others, the G4 subclass comprises the Polysomy 7 and Interferon subgroups of the Chiang classification. Subgroups G5 and G6 are characterized by mutations in the *CTNNB1* gene leading to activation of canonical Wnt/β-catenin signaling. These HCCs frequently show a pseudo-glandular (cholestatic) growth pattern. [50, 56] Detection of nuclear β-catenin and diffuse glutamine synthetase expression can be used to identify HCCs belonging to Boyault´s subclasses G5/6. The activation of Wnt/β-catenin signaling pathway is stronger in G6 compared to the G5 subgroup. Mutations in the promoter of the *TERT* gene and methylation-induced inactivation of *CDKN2A* and *CDH1* are listed among the molecular characteristics of these HCC subclasses. [50, 59]
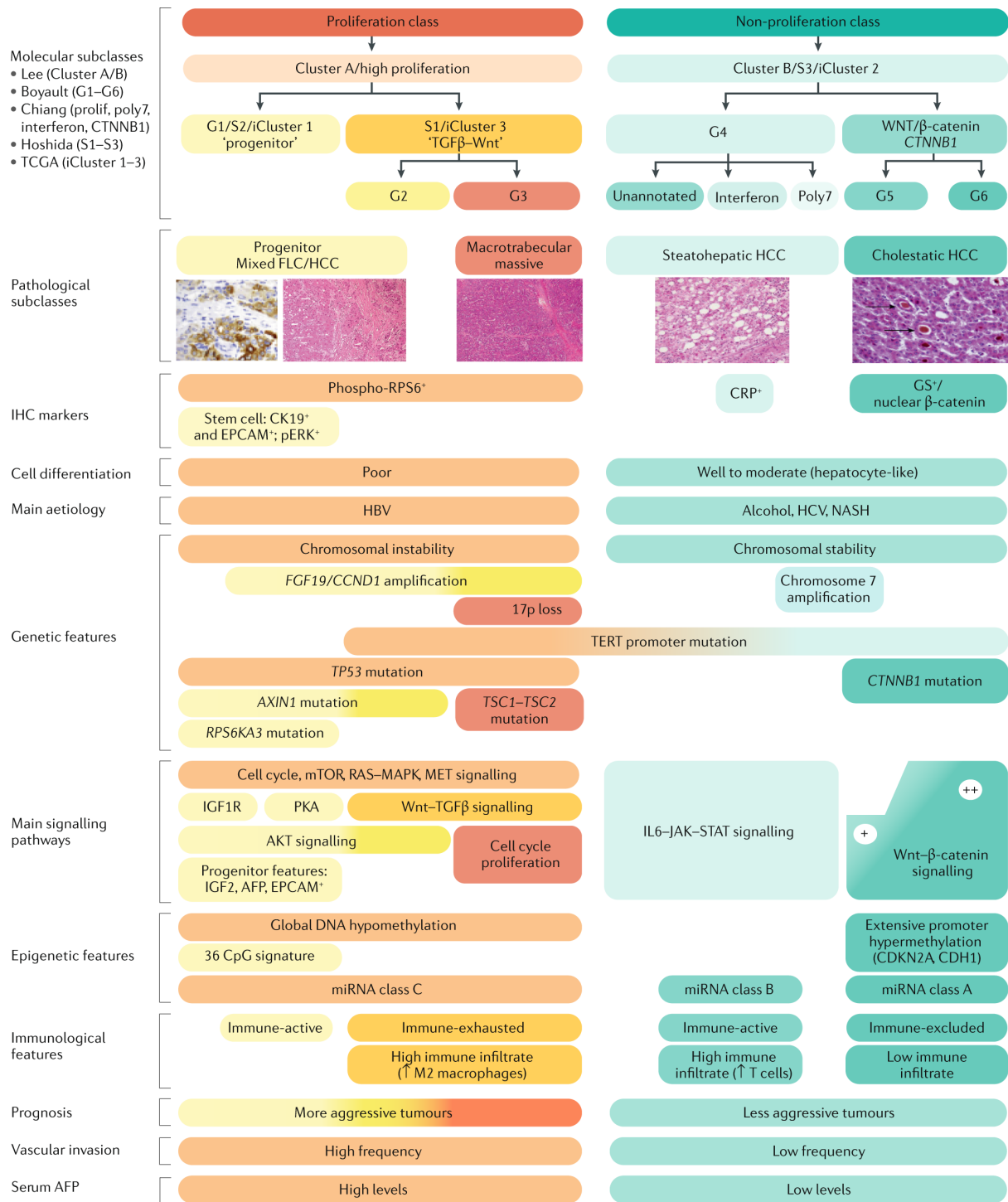
Figure 3. Schematic summary of molecular classification of HCC. [1]

1.3.2 Cancer driver gene mutations

High throughput next-generation exome sequencing enabled the identification of cancer driver genes frequently altered in HCC. These either activate oncogenes or inactivate tumor suppressor genes. The most frequent somatic mutations observed in about 60% of HCC

cases affects the *TERT* gene promoter. [47] Mutations in *CTNNB1* (β-catenin encoding gene) and inactivation of *AXIN1*, *AXIN2* or *APC* resulting in activation of the Wnt–β-catenin signaling pathway have been described in 54% of the HCC cases. [60] Genetic alterations in genes associated with cell cycle control like *TP53*, *RB1*, *IRF2*, *CDKN2A/B*, *CCNA2*, *CCND1*, *CCNE1*, and *CREBBP* were present in 49% of HCC. Mutations in genes participating in chromatin remodeling (e.g., *ARID1A* and *ARID2*), epigenetic regulation (e.g., *KMT2D*), oxidative and ER stress pathways (e.g., *NFE2L2*), the PI3K-AKT-mTOR, and MAPK pathways (e.g., *RPS6KA3*, *KRAS*, *PIK3CA*) were observed in at least one third of HCC respectively. [60, 61] Chromosomal alterations are recurrently found in human HCC. Losses of genetic information affect the chromosome arms 1p, 4q, 6q, 8p, 9p, 13q, 16p, 16q and 17p, while chromosomal gains can be detected at 1q, 5p, 6p, 7p, 7q, 8q, 17q, 20p, 20q and Xq. [62] Importantly, regions showing gained genetic information harbor known oncogenes like *CCND1/FGF19*, *MDM4*, *MET*, *MYC*, and *VEGFA*. These amplifications lead to upregulation of oncogene expression levels and consequently activation of associated oncogenic signaling pathways. [48, 51, 63]

1.3.3 Mosaic mouse models

The development and introduction of genetically engineered mouse models (GEMM) led to substantial advances in liver cancer research. For instance, it allowed for the identification and validation of cancer related genes and investigations on their roles during hepatocarcinogenesis. [64] Direct microinjection of recombinant DNA into a fertilized murine egg or lentiviral transduction of embryonic stem cells are conventional genetic manipulations used to generate transgenic mouse models of liver cancer. [65, 66] These mouse models harbor the cancer initiating mutation in all cells of their body and therefore are valuable tools to study familial cancer syndromes. These constructs may carry inducible and/or tissue specific promoters facilitating spatiotemporal control of transgene expression. This approach allows to investigate the role of the studied gene in different stages of tumor development and progression. [66] However, the use of tissue-specific promoters generates a bias towards certain epithelial cells and does not mimic the situation of spontaneous tumorigenesis in humans. In addition, the utilized promoters represent the minimum sequence essential for detection and binding of transcription factors and do not thoroughly resemble the endogenous promoter. In addition, they lack cis- and trans-regulatory elements. Importantly, this experimental approach is also cost- and time-intensive and may not be feasible at all due to embryonic lethality or developmental compensation during tissue growth and maturation. [64,

66] Thus, more sophisticated animal models, such as mosaic cancer mouse models or the use of transposons for mutagenesis screens, have been developed and been instrumental for the discovery of novel tumor driver genes. [64, 67, 68]

First introduced by Benedito and colleagues, mosaic mouse models resemble the intra- and inter-tumor heterogeneity and are valuable tools to study gene functions during clonal evolution of cancer. In mosaic models of cancer, tumor cells are the only cells harboring the studied genetic aberration and their surrounding consists of non-mutated parenchyma cells. Therefore, development and growth of these tumors in the context of a genetically wild-type microenvironment better replicates the sporadic process of human hepatocarcinogenesis. [69, 70] One way to generate chimeric mouse models of liver cancer includes *ex vivo* genetic modification and expansion of hepatic stem/progenitor cells and re-transplantation of successfully transformed hepatoma cells into the liver of recipient mice carrying the desired genetic background. [71]

The alternative method to generate mosaic mouse models of liver cancer is the delivery of plasmid DNA into mature hepatocytes via hydrodynamic tail vein injection (HTVI). This approach is more simple and less expensive compared to recombination methods. [72] The unique anatomical structure of hepatic sinusoids and the close anatomical association of hepatocytes and fenestrated sinusoidal epithelium makes liver the ideal target for HTVI. For this approach, large amount of naked plasmid DNA (about 50 µg) is dissolved in saline solution (equivalent to 8-10% of the mouse body weight). Rapid injection (5 to 9 seconds) of such high volume into the murine lateral tail vein generates the hydrodynamic force to expand the inferior vena cava and overstretch the myocardial fibers inducing a temporary acute cardiac insufficiency leading to hepatic congestion and the influx of the DNA solution into the liver. The hydrodynamic pressure in the endothelial barrier and leads to formation of pores in the plasma membrane of the surrounding hepatocytes driving the intracellular uptake of the injected plasmid DNA. In less than two minutes, the pores in plasma membrane become closed again and the injected genetic material remains in hepatocytes interior. [73-75] The transfection efficacy of HTVI in liver is 10 to 40% of all hepatocytes and in other organs including kidney, spleen, lung, and heart is <0.1%. Thus, HTVI specifically targets liver and due to its partial transfection efficacy leads to generation of mosaic mouse models of liver cancer. [76]

However, liver tumor formation requires continuous expression of the oncogenic DNA, but the DNA delivery to hepatocytes is not sufficient to induce liver cancer *per se* as the

exogeneous genetic material would be degraded in these cells. To achieve long-term stable expression of the delivered DNA, the so-called cut and paste function of the sleeping beauty (SB) transposases can be combined with HTVI-mediated delivery of target genes. This genetic tool consists of two components: the SB transposase and the transposon element which is flanked by inverted repeats. The SB transposase recognizes and binds to inverted repeats at which it precisely excises the flanked DNA sequence and reintegrates it into a TA dinucleotide-rich region within the host genome promoting constitutive gene expression. [77, 78]

The multistep process of hepatocarcinogenesis suggests that accumulation of multiple cancer driver gene mutations is required for the induction of liver cancer. Congruently, single gene oncogenic mutations have been shown to be insufficient to induce liver cancer in genetically engineered mouse models. Interestingly, combined expression of cancer driver genes is able to lower the latency period of tumorigenesis and to increase the tumor incidence up to 100%. [79] For instance, MYC is a transcription factor regulating expression of various genes involved in proliferation and cell growth; its overexpression contributes to the pathogenesis of several malignant tumors. [80] MYC overexpression is present in up to 70% of primary HCCs. [63] Nevertheless, overexpression of MYC alone is insufficient to induce HCC tumors in transgenic and mosaic mouse models and requires co-expression of another cancer driver like transforming growth factor α (TGF-α), transcription factor E2F1 (E2F1) or AKT. [79, 81-83] For instance, HTVI-mediated transfection of murine hepatocytes with the transposon vector encoding both murine MYC and activated (myristoylated) human AKT1 markedly accelerates tumor formation in mosaic mouse livers (4 weeks after injection). [68]

Functional validation of tumor suppressor gene candidates is essential to discriminate relevant tumor suppressors from passenger genes. To address this issue, Zender et al. have technically integrated cancer genomics, *in vivo* RNA interference (RNAi) technology, and HTVI to develop a rapid and cost-effective screening method for the discovery and the validation of tumor suppressor genes in hepatocellular carcinoma. In the initial approach, immortalized p53-deficient and MYC overexpressing embryonic liver progenitor cells were infected with an shRNA library pool. This cell population was then transplanted into nude mice either subcutaneously or by intrasplenic injection into the liver and tumor development was monitored. [67] Lately, HTVI of transposon vectors has replaced the *ex vivo* manipulation and re-injection and has become the state-of-the-art technology for oncogene and shRNA library delivery into the murine liver. [84, 85]

1.4 Serine Proteases

1.4.1 Serine protease protein family

Serine proteases account for almost one third of all proteases. They are characterized by the nucleophilic serine residue in their active site and are able to cleave peptide bonds. The active site of serine proteases mainly consists of Asp-His-Ser residues, the so-called catalytic triad. [86] According to MEROPS database, serine proteases are classified into 13 clans and 40 families. Four clans of serine proteases represented by chymotrypsin, subtilisin, carboxypeptidase Y, and caseinolytic protease (CLPP) contain the classic Asp-His-Ser catalytic triad. [87, 88] The other clans share the serine residue but contain different triads or dyads. The chymotrypsin-like serine proteases are the best studied subgroup. They take part in a wide variety of physiological processes including digestion, hemostasis, apoptosis, signal transduction, reproduction, and immune response, to name a few. Of note, the classical configuration of the catalytic triad has successfully performed over distinct evolutionary realms and is thus widely found in eukaryotes, prokaryotes, archaea, and viruses. Therefore, the specificity required for alternative physiological roles might have lied in the topology of the substrate binding sites (termed the active site cleft) leading to discrimination in substrate selectivity, or in the rate of chemical transformation. [86]

The protein structure of chymotrypsin-like serine proteases comprises domains for substrate recognition, zymogen activation, and catalysis. [86] The catalytic component consists of the catalytic triad and the oxyanion hole. [86, 88] The oxyanion hole is a positively charged pocket located in the active site. It mediates catalysis via stabilizing transition state of negative charge lowering the activation energy required for execution of the catalytic reaction. [89] The substrate recognition component is called the S1 site which denotes a pocket adjacent to Ser195 formed by the residues 189-192, 214-216, and 224-228. The specificity of the trypsin-like serine proteases is specified in terms of the P1-S1 interaction and the hydrophobicity of the residues in S1 site determines this specificity for different substrates. Majority of trypsin-like serine proteases are synthesized in the form of an inactive zymogen precursor. Proteolytic processing of the proprotein precursor is required for the activation of the enzyme and is basically the cleavage between residues 14 and 15 of N-terminus segment which releases the N-terminal Ile16. The emerging N-terminus forms the Ile16-Asp194 salt bridge inducing conformational changes in the protein. This reforms the activation domain (the N-terminus to residue 19, residues 142-152, 184-193, and 216-223) and leads to formation of S1 site and

oxyanion hole, thereby the active protease. The Ile16-Asp194 salt bridge and oxyanion hole are structurally linked via the Ser195 of the catalytic triad. [86]

The catalytic mechanism of action of serine proteases is very well studied. Serine proteases cleave polypeptides at the location of their peptide bond. Binding of the substrate to the S1 site of the protease induces a slight conformational switch in the enzyme bringing the side chain of the catalytic Asp102 in closer proximity to the imidazole ring of the catalytic His57. The negative charge of the oxygen in the carboxylic acid group of the Asp102 modifies the electronic configuration of the histidine ring. As a result, the nitrogen of the histidine ring withdraws the proton from the side chain of the Ser195 creating an alkoxide ion (negatively charged oxygen). This exceptionally reactive alkoxide ion that is striving for a nucleophile, then attacks on carbonyl carbon of the peptide bond in the substrate. This rearranges the carboxyl group in the peptide bond and forms a tetrahedral intermediate which is chemically unstable. The oxyanion hole of the enzyme mediates the stabilization of the tetrahedral intermediate. Meanwhile the nitrogen of the amino group abstracts the hydrogen from the histidine ring which breaks apart the peptide bond and releases half of the substrate. It follows that the nitrogen of the histidine ring that has lost the hydrogen attacks a molecule of water and takes a proton from it. The remaining hydroxide makes a nucleophilic attack on the carbonyl carbon and forms a second tetrahedral intermediate. In order to restabilize the tetrahedral intermediate, the oxygen of carboxyl group captures the proton from the histidine ring and breaks its bond with the Ser195. Doing so the other half of the substrate is released and the enzyme returns to its initial state. [86, 88, 90]

1.4.2 Serine proteases in hepatocellular carcinoma

Proteolytic cleavage is required for proper folding and functional activation of proteins. The dysregulation of serine proteases has been implicated in the pathogenesis of numerous cancer types. Given the diverse physiological functions of serine proteases in various intra- and intercellular processes, serine proteases have been primarily implicated in carcinogenesis via remodeling of the extracellular matrix and as a driver of cancer cell migration and invasion. Nonetheless, post-translational processing of growth factors, cytokines, and proangiogenic factors as well as modulating cell-cell and cell-matrix associated molecules are listed among other functions of serine proteases. Their dysregulation may thus confer a growth advantage to neoplastic cells. [91, 92] As an example, the upregulation of type II transmembrane serine proteases like Hepsin, TMPRSS2, and TMPRSS4 leads to increased invasion and metastasis

in several cancers. [92] In hepatocytes, Hepsin mediates the proteolytic activation of the pro-HGF/Met signaling and regulates downstream glycogen, lipid, and protein metabolism via AKT-mTOR signaling. [93] Likewise, Partanen et al. proposed that the mislocalization of desmosomal Hepsin to the cytoplasm in hepatocytes induces distortion of basal membrane integrity thereby facilitating MYC-driven tumorigenesis. [94] TMPRSS4 has been reported to promote epithelial-to-mesenchymal transition (EMT), angiogenesis, and tumor progression in human HCC. [95] Similarly, the secretory High-temperature requirement A serine peptidase 1 (HTRA1) is associated with EMT and tumor migration in HCC. [96] Mitochondrial serine proteases such as mitochondrial ATP-dependent Clp protease proteolytic subunit (CLPP), mitochondrial Lon protease homolog (LONP1), and HTRA2 are chaperones involved in maintaining the protein quality in mitochondria and thus protect the cells against oxidative stress. [97] In addition, it is reported that overexpression of CLPP restores mitochondrial bioenergetics and respiration following stress conditions and inhibits the development of late-stage senescence in HCC cells. [98] Likewise, LONP1 maintains mitochondrial and cellular homeostasis and its dysregulation may promote tumor progression. [99] Taken together, serine proteases have been implicated in the pathogenesis of cancer and may act either in the intracellular or in the extracellular environment.

1.4.3 Serine protease 23

Serine protease 23 (PRSS23, SIG13, SPUVE) is an evolutionarily highly conserved protease in vertebrates. It belongs to the clan PA and the family S1 (chymotrypsin A) of peptidases. [87] In contrast to the S1B family mediating catalysis of intracellular proteins, the S1A family is primarily involved in extracellular cleavage. [88] The gene encoding PRSS23 protein is located on chromosome 11q14.2 (containing 8 exons) in humans and on chromosome 7 (containing 7 exons) in mice (*mus musculus*). [100] In humans, PRSS23 is ubiquitously expressed [101] and 9 splice variants have been reported for *PRSS23* gene. [102] The protein sequence is consisted of a signal peptide spanning the amino acids 1-19 and a polypeptide chain (amino acids 20 -383), which forms the mature protein after proteolytic processing. In terms of post-translational modifications, two glycosylation sites at amino acid residues 93 and 207 and one phosphorylation site at residue 109 are reported. Moreover, one disulfide bond between residues 160 and 176 of PRSS23 is predicted by similarity analysis. The molecular weight of the longest isoform is 43 kDa. [101]

Initially, *Prss23* was identified in mice as an ovarian protease with high homology to *Prss35*. The spatiotemporal expression pattern of *Prss23* during the murine ovarian cycle suggested that PRSS23 plays a key role in maturation of antral follicles probably via remodeling of ECM and/or regulating the availability of insulin-like growth factor. [103] In another study, *Prss23* was detected in the ovarian stroma and theca tissues just before ovulation and high levels were detected in atretic ovarian follicles. [104] Moreover, PRSS23 has been implicated in cardiogenesis in humans and mice. [105] There is evidence that PRSS23 drives initiation of endothelial-to-mesenchymal transition (EndoMT) and plays a pivotal role in proper development of the atrioventricular canal in zebra fish, potentially via modulation of TGF-β-induced Snail transcription. [106] In addition, various roles including mediating collagen type I degradation [107] and bone metabolism [108] have been implicated for PRSS23.

Expression profiling revealed upregulation of PRSS23 expression in various cancer entities including breast [109], prostate [110], pancreatic [111, 112], papillary thyroid [113], and gastric cancer [114]. However, the functional role of PRSS23 in cancer remains largely elusive. Chan et al. reported a higher expression level of PRSS23 in ERα-positive breast cancers compared to ERα-negative counterparts, which was attributed to direct binding of estrogen receptor α (ERα) to the upstream promoter region (2029 – 342 bp) of *PRSS23*. In addition, PRSS23 might play a critical role in estrogen-induced proliferation of breast cancer cells. [115] In addition, Han et al. demonstrated an association of PRSS23 expression levels with tumorigenesis and progression of human gastric cancer. In this cancer entity, the inhibition and knockdown of PRSS23 respectively decreased cell viability and inhibited tumor growth in an EIF2 signaling dependent manner. Notably, the PRSS23 knockdown was associated with downregulation of MYC and AKT1 and led to inhibition of EIF2 signaling. [114]

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Consumables

Table 1. list of the consumables used to perform the described experiments of this study.

Item	Catalogue No.	Supplier
1000 µl XL graduated TipOne® Filter Tip, Natural, Racks (sterile)	S1122-1830	STARLAB
100 µl Bevelled TipOne® Filter Tip, Natural, Racked (sterile)	S1120-1840	STARLAB
25 cm² Cell Culture OrFlask with vented cap	4410100N	Orange Scientific
Amersham™ Protran® Western blotting membranes, nitrocellulose	GE10600002	Amersham
Amicon Ultra-15, PLGC Ultracel-PL Membran, 10 kDa	UFC901008	Merck Millipore
B Braun Sterican Needles 0.90 × 70mm Sterile	#4665791	B. Braun
BD Discardit™ II syringe (20 mL)	#300296	BD
BD Microlance™ 3 Needles	#302200	BD
BlueStar PLUS Prestained Protein Marker	MWP04	NIPPON Genetics
Brand Reagent reservoir 60 ml for multichannel pipettes (sterile)	#7-4621	neoLab
Centrifuge tubes, 15 mL	GN188271	Greiner
Centrifuge tubes, 50 mL	GN227261	Greiner
Corning® BioCoat™ Growth Factor Reduced Matrigel® Invasion Chamber	#354483	Corning
Cover Glasses round	#01-0012/1	R. Langenbrinck
Cryogenic vials	GN122279	Greiner
DISTRITIP Maxi ST, 12.5 mL	F164150	Gilson
DISTRITIP Micro ST, 125 µL	F164130	Gilson
DISTRITIP Mini ST, 1250 µL	F164140	Gilson
Eppendorf Tube®, 5 mL	EP0030119460	Eppendorf
Falcon® 96-well Clear Round Bottom Assay Microplate	#353910	Corning
Falcon® Cell culture 12-well plate	#353043	Corning
Falcon® Cell culture 24-well plate	#353047	Corning
Falcon® Cell culture 48-well plate	#353078	Corning

Falcon® Cell culture 6-well plate	#353046	Corning
Falcon® Cell culture 96-well plate	#353077	Corning
Greiner Bio-One CELLSTAR™ TC treated Cell Culture Dish with vents	#664160	Greiner
Greiner Bio-One™ 1000µL Pipette Tips	#740290	Greiner Bio-One
MicroAmp™ Fast Optical 96-Well Reaction Plate, 0.1 mL	#4346907	Thermo Fisher Scientific
MicroAmp™ Optical Adhesive Film	#4311971	Thermo Fisher Scientific
Microscope slides, Menzel Gläser 76×26 mm	#630-1985	Thermo Scientific
neoCulture Cell scrapers ABS	C-8222	neoLab
Odyssey® Nitrocellulose Membranes	#926-31092	LI-COR
PARAFILM® M	PM-996	Bemis Mfg
Petri dish 92×16mm with cams	SAR-821473001	Sarstedt
Pipette tip 200 µL	#70760002	Sarstedt
Quali - PCR - Tubes 0.2 ml with individually attached cap, strip of 8-1×120 Str., flat cap	G003-SF	Kisker Biotech
Falcon® round bottom 14 ml test tubes	COR-352006	Corning
Safe-Lock Tubes 1.5 mL	EP0030120086	Eppendorf
Safe-Lock Tubes 2.0 mL	EP0030120094	Eppendorf
SafeSeal Tips Professional 10 µl, Sterile	#693010	Biozym
Seahorse XFe96 FluxPak mini	#102601-100	Agilent
Sodium Pyruvate (100 mM)	#11360-039	Gibco
Solid-glass beads	Z265926-1EA	Sigma-Aldrich
Stripette™ Serological Pipets, 10 mL	S058.4488	Corning
Stripette™ Serological Pipets, 25 mL	S058.4489	Corning
Stripette™ Serological Pipets, 5 mL	S058.4487	Corning
Sterile syringe filters, pore size 0.45 µm cellulose acetate membrane	#28145-481	VWR
T4 DNA Ligase Reaction Buffer (10×)	B0202S	New England Biolabs
Whatman 3 CHR	#3003-917	Cytiva

2.1.2   Chemicals, drugs, and antibiotics

Table 2. List of the chemicals, drugs, and antibiotics used in the described experiments of this study.

Chemical/drug/antibiotic	Catalogue No.	Supplier
2-Mercaptoethanol	A1108	AppliChem
Acetic acid	#3738.1	CARL ROTH
Agarose	A9539	Sigma-Aldrich
Albumin Fraction V, biotin-free, NZ-Origin	#163.4	CARL ROTH
Ammonium peroxydisulphate	#9592.2	CARL ROTH
Ampicillin, Ready Made Solution (100 mg/mL)	A5354	Sigma-Aldrich
Bacto™ Agar	#214010	BD
Blasticidin S HCl (10 mg/mL)	A1113903	Gibco
Boric acid	LC-5914.1	neoFroxx GmbH
Bromophenol Blue	#B8026	Sigma-Aldrich
Cell Lysis Buffer (10×)	#9803	Cell Signaling
Coomassie® Brilliant Blue G 250	#17524	SERVA Electrophoresis
Crystal Violet	C0775	Sigma-Aldrich
Dimethyloxaloylglycine (DMOG)	sc-200755	Santa Cruz Biotechnology
Dimethylsulfoxid (DMSO)	A994.2	CARL ROTH
DTT	#39759	SERVA Electrophoresis
EDTA	#34549	Fluka Analytical
Ethanol	#32205	Honeywell Riedel-de Haën™
Gammabind Plus Sepharose®	GE17-0886-01	Sigma-Aldrich
Glycerol	#3783.1	CARL ROTH
Glycine	LC-4522.2	neoFroxx GmbH
Hexadimethrine bromide (Polybrene)	#107689	Sigma-Aldrich
IGEPAL® CA-630 *	I8896	Sigma-Aldrich
Kanamycin A	K4000	Sigma-Aldrich
LB Agar (Luria/Miller)	X969.1	CARL ROTH
LB Broth (Luria/Miller), granulated	#6673.1	CARL ROTH
Methanol	#20847.307	VWR International S.A.S.
Penicillin-Streptomycin	#15140122	Gibco
Phenylmethylsulfonyl fluoride (PMSF)	P7626	Sigma-Aldrich

Polyethylenimine (PEI)	#23966	Polysciences
Powdered milk	T145.2	CARL ROTH
Protein A PLUS-Agarose	sc-2001	Santa Cruz Biotechnology
Protein A/G PLUS-Agarose	sc-2003	Santa Cruz Biotechnology
Protein G PLUS-Agarose	sc-2002	Santa Cruz Biotechnology
Puromycin Dihydrochloride	A1113803	Gibco
ROTI®Stock 20% SDS	1057.1	CARL ROTH
ROTIPHORESE®Gel 30 (37.5:1)	#3029.1	CARL ROTH
SDS	#20765	SERVA Electrophoresis
Sodium chloride (NaCl)	LC-5932.3	neoFroxx GmbH
TEMED	#2367.3	CARL ROTH
Thiazolyl Blue Tetrazolium Bromide (MTT)	M2128	Sigma-Aldrich
Tris	#4855.2	CARL ROTH
Tween® 20	#9127.1	CARL ROTH

* Substitute of NP-40

2.1.3 Commercially-available reagents and kits

Table 3. List of the reagents and kits used to perform the described experiments of this study.

Reagent/Kit	Catalogue No.	Supplier
Agilent Seahorse XF Real-time ATP Rate Assay kit	#103677-100	Agilent
Bradford Reagent	B6916	Sigma-Aldrich
BlueStar PLUS Prestained Protein Marker	MWP04	NIPPON Genetics
dNTP Mix (10 mM)	R0192	Thermo Fisher Scientific
Dulbecco's Phosphate-Buffered Saline (D-PBS), no calcium, no magnesium	#14190169	Gibco
Dulbecco's Modified Eagle Medium (DMEM), high glucose	#41965062	Gibco
EZ-Run™ Prestained Rec Protein Ladder	BP36031	Fisher BioReagents
Fetal Bovine Serum	#10270106	Gibco
FxCycle™ PI/RNase Staining Solution	F10797	Thermo Fisher Scientific
Gel Loading Dye, Purple (6×)	B7024S	New England Biolabs

GeneChip® Hybridization, Wash and Stain Kit	#900720	Affymetrix
GeneChip™ Mouse Gene 2.0 ST Array	#902500	Affymetrix
GeneChip® WT Plus Reagent Kit	#902280	Affymetrix
Glucose Solution	#A2494001	Gibco
L-Glutamine (200 mM) (100×)	#25030081	Gibco
MEM Non-Essential Amino Acids Solution (100×)	#11140-035	Gibco
Mitochondria Isolation Kit for Cultured Cells	#89874	Thermo Fisher Scientific
Mix & Go E. coli Transformation Kit	T3001	Zymo Research
Monarch® Plasmid Miniprep Kit	T1010	New England Biolabs
Mouse Serine Protease 23 (PRSS23) ELISA Kit	MBS9717245	MyBioSource
MycoAlert™ Mycoplasma Detection Kit	LT07-418	Lonza
NEBuilder® HiFi DNA Assembly Master Mix	E2621	New England Biolabs
NE-PER™ Nuclear and Cytoplasmic Extraction Reagents	#78833	Thermo Fisher Scientific
Nuclear Extract Kit	#40010	Active Motif
NucleoSpin RNA	#740955250	Macherey-Nagel
Opti-MEM® I Reduced-Serum Medium	#31985070	Gibco
Poly-D-Lysine	A3890401	Gibco
PowerUp™ SYBR™ Green Master Mix	A25777	Thermo Fisher Scientific
PureYield™ Plasmid Midiprep System	A2495	Promega
Quick-Load® 1 kb Extend DNA Ladder	N3239	New England Biolabs
Quick Western Kit – IRDye® 680RD	#926-68100	LI-COR
QuickExtract™ DNA Extraction Solution	QE09050	Lucigen
QuikChange Lightning Site-Directed Mutagenesis Kit	#210519	Agilent
ReliaPrep™ DNA Clean-Up and Concentration System	A2893	Promega
Seahorse XF DMEM medium, pH 7.4	#103575-100	Agilent
SureSelect Human All Exon v2	G3353A	Agilent Technologies
TransAM® c-Myc	#43396	Active Motif
Trypsin-EDTA (0.05%), phenol red	#25300096	Gibco

UltraPure™ DNase/RNase-Free Distilled Water	#10977035	Invitrogen
Wizard® SV Gel and PCR Clean-Up System	A9282	Promega

2.1.4 Enzymes

Table 4. List of the enzymes used in performing the described experiments of this study.

Enzyme	Catalogue No.	Supplier
BamHI-HF	R3136	New England Biolabs
EcoRI-HF	R3101	New England Biolabs
Gateway BP Clonase II Enzyme mix	#11789020	Thermo Fisher Scientific
Gateway LR Clonase II Enzyme mix	#11791020	Thermo Fisher Scientific
NotI-HF	R3189	New England Biolabs
PhosStop	#4906837001	Roche
Phusion High-Fidelity DNA Polymerase	F530L	Thermo Fisher Scientific
Protease-Inhibitor Mix G	#39101.03	SERVA Electrophoresis
rDNase	CAS9003-98-9	MACHEREY-NAGEL
RevertAid H Minus Reverse Transcriptase	EP0452	Thermo Fisher Scientific
SalI-HF	R3138	New England Biolabs
T4 DNA Ligase	M0202	New England Biolabs

2.1.5 Antibodies

Table 5. List of the primary and secondary antibodies used for Western blotting and immunoprecipitation experiments described in this study.

Target protein	Catalogue No.	Supplier
ACTIN	#A5441	Sigma-Aldrich
ANAPC1	NB100-86985	Novus Biologicals
c-MYC	ab32072	Abcam
COX-2	#610204	BD Transduction Laboratories
DYKDDDDK Tag (FLAG)	#2368	Cell Signaling
EMILIN-1	sc-365737	Santa Cruz Biotechnology
FLAG	F1804	Sigma-Aldrich
GAPDH	MCA4739	Bio-Rad
GMPS	sc-376163	Santa Cruz Biotechnology

IRDye 800CW Goat Anti-Mouse	#926-32210	LI-COR
IRDye 800CW Goat Anti-Rabbit	#926-32211	LI-COR
IRDye 800CW Goat Anti-Rat	#926-32219	LI-COR
IRDye 680RD Donkey anti-Mouse IgG	#925-68072	LI-COR
IRDye 680RD Goat anti-Rabbit IgG	#925-68071	LI-COR
Keratin 19	ab133496	Abcam
MCM3	S4012	Cell Signaling
MCM6	ab201683	Abcam
Normal mouse IgG	sc-2025	Santa Cruz Biotechnology
OGT	#24083	Cell Signaling
PARP	#9542	Cell Signaling
Peripherin	NB300-137	Novus Biologicals
phospho-AKT (Ser473)	#4060	Cell Signaling
phospho-AKT (Thr308)	#13038	Cell Signaling
phospho-AMPKα (Thr172)	#2535	Cell Signaling
phospho-S6 ribosomal protein (Ser235/236)	#2211	Cell Signaling
PLD1	#3832	Cell Signaling
p-mTOR	sc-293133	Santa Cruz Biotechnology
PRSS23	NBP1-79488	Novus Biologicals
	LS-C373838	LifeSpan BioSciences

2.1.6 Primers

Table 6. List of the primer pairs used for PCR and qRT-PCR experiments described in this study.

Primer application	Primer sequence Forward 5′ to 3′	Primer sequence Reverse 5′ to 3′
QuikChange® primer (c.686C>G)	CGCACCCATGTGGCCAAGGG GTGGA	TCCACCCCTTGGCCACATGG GTGCG
PCR amplification of *Prss23* for BP reaction	GGGGACAAGTTTGTACAAAA AAGCAGGCTCCATGGCTGGA ATCCCGGGGCT	GGGGACCACTTTGTACAAG AAAGCTGGGTTCCCCTCCCT GCAATCTAGG
PCR amplification of *Prss23*-FLAG C for HiFi DNA assembly	TCCATTTCAGGTGTCGTGAGA TGGCTGGAATCCCGGGG	TAGAGCGGCCGCCCTCGAG GTCACTTGTCGTCATCGTCT TTG
Sequencing of insert (*Prss23* (-FLAG C))	TCAAGCCTCAGACAGTGGTTC	TATAGACAAACGCACACCG

in pLV-EF1a-IRES-Blast		
Sequencing of insert (*Prss23*) in pDONR201	TCGCGTTAACGCTAGCATGGATCTC	GTAACATCAGAGATTTTGAGACAC
Sequencing of *Prss23*-FLAG C in pDEST26-FLAG-C-Puro	CGC AAA TGG GCG GTA GGC GTG	TACGCCCTGCTGGAACTCAAGAAA
Sequencing of insert (*Prss23*) in pCaggs	TGTGACCGGCGGCTCTAGA	AATCCCATACAACCCTGAGTCCC
Sequencing *Prss23*		CGGGCCGGTCATTGTCATA
Sequencing *Prss23*	GAGATCTCGCAGGAAGAGGC	
Sequencing *Prss23*	TACGCCCTGCTGGAACTCAAGAAA	
Sequencing *Prss23*		TGGTGCTGGCAAGAAGACG
Sequencing *Prss23*		TTGCCTATTGAAGACTTTGCTATC
Sequencing *Prss23*		AAACCTGCCATCGTAGCCATAAAT
Sequencing *Prss23*		TGAGGCAAGACTACAGGAGGCG
qRT-PCR of exogenous *Prss23*	TCGACGCCAAAGCGAAATTG	GTACTGCTTGGCCTCTTCGT
qRT-PCR of *Gapdh*	TGTCCGTCGTGGATCTGAC	CCTGCTTCACCACCTTCTTG

All sequences were synthesized by Microsynth (Switzerland). Lyophilized primers were dissolved in nuclease-free water to a final concentration of 100 μmol/L. Primers for qRT-PCR were designed using the Primer-BLAST tool from the National Center for Biotechnology Information (NCBI). [115] The primers were spanning an exon-exon junction and targeting a sequence sized 70-150 bp.

2.1.7 shRNA oligos

Table 7. List of the shRNA oligos used in *in vivo* shRNA library screening of this study.

shRNA-ID	Strand	Sequence 5′ to 3′
Nmnat2.2587	Sense	ACGGGTTCTTATTTGAAAACAA
	Antisense	TTGTTTTCAAATAAGAACCCGG
Nmnat2.3000	Sense	ATGGGCAGATTGTATCTTGTTA
	Antisense	TAACAAGATACAATCTGCCCAC
Nmnat2.3296	Sense	CTGGCACAGAGATGGTAATAAA
	Antisense	TTTATTACCATCTCTGTGCCAT
Mll2.10420	Sense	CCCCAGACACAGACCTAGACAA

	Antisense	TTGTCTAGGTCTGTGTCTGGGA
Mll2.18506	Sense	ACCGGGTTGTATAGTTGGACTA
	Antisense	TAGTCCAACTATACAACCCGGG
Mll2.3314	Sense	ACACGAGATGGAGACTGATAAA
	Antisense	TTTATCAGTCTCCATCTCGTGC
Hoga1.1285	Sense	AAAGCAGAGAGTTCTACATCTA
	Antisense	TAGATGTAGAACTCTCTGCTTC
Hoga1.1288	Sense	ACAGAGAGTTCTACATCTAGAT
	Antisense	ATCTAGATGTAGAACTCTCTGC
Hoga1.1291	Sense	AAGAGTTCTACATCTAGATCTA
	Antisense	TAGATCTAGATGTAGAACTCTC
Dlk1.1850	Sense	AAACGATGGATATATACCAAAA
	Antisense	TTTTGGTATATATCCATCGTTC
Dlk1.3623	Sense	CTCCAACGTGACCAAAGATCAA
	Antisense	TTGATCTTTGGTCACGTTGGAA
Dlk1.364	Sense	ACCCTCTGTGACAAGTGTGTAA
	Antisense	TTACACACTTGTCACAGAGGGG
Dip2c.1236	Sense	ACAGCTTGTCAATACCCTTAAA
	Antisense	TTTAAGGGTATTGACAAGCTGC
Dip2c.1539	Sense	AAAGGTTGCTTACAACATTCTA
	Antisense	TAGAATGTTGTAAGCAACCTTC
Dip2c.3656	Sense	AAACGACATTGCCTACAGTCAA
	Antisense	TTGACTGTAGGCAATGTCGTTG
Clmp.1160	Sense	CACGAGGAAGAAGACAGACCTA
	Antisense	TAGGTCTGTCTTCTTCCTCGTA
Clmp.2233	Sense	AAGGCTTGACAACAGATATCAA
	Antisense	TTGATATCTGTTGTCAAGCCTC
Clmp.2241	Sense	CCAACAGATATCAAGACGATTA
	Antisense	TAATCGTCTTGATATCTGTTGT
Kcnq2.423	Sense	ACACGCCTACGTGTTCCTTTTA
	Antisense	TAAAAGGAACACGTAGGCGTGG
Kcnq2.476	Sense	ACACCATCAAGGAGTACGAGAA
	Antisense	TTCTCGTACTCCTTGATGGTGG
Kcnq2.547	Sense	CTCGGTGTTGAGTACTTTGTGA
	Antisense	TCACAAAGTACTCAACACCGAA

Map3k4.1209	Sense	CCACGAAAGATCTAAATCAGAA
	Antisense	TTCTGATTTAGATCTTTCGTGA
Map3k4.3180	Sense	CCAGGCAGAAGATAGGAGACAA
	Antisense	TTGTCTCCTATCTTCTGCCTGA
Map3k4.339	Sense	ACCGATCGAACTTGAAAGAAAA
	Antisense	TTTTCTTTCAAGTTCGATCGGC
Rtn4ip1.1283	Sense	CAACGTTGGTGGATCTACTGAA
	Antisense	TTCAGTAGATCCACCAACGTTA
Rtn4ip1.1618	Sense	ACCGAGGAAAGACTGTGGTTAA
	Antisense	TTAACCACAGTCTTTCCTCGGG
Rtn4ip1.1628	Sense	AACTGTGGTTAATGTAGTTTAA
	Antisense	TTAAACTACATTAACCACAGTC
Prss23.1023	Sense	ATACCGCTTCTGTGATGTCAAA
	Antisense	TTTGACATCACAGAAGCGGTAC
Prss23.1249	Sense	ACCCAGATTTGCTATTGGATTA
	Antisense	TAATCCAATAGCAAATCTGGGC
Prss23.2362	Sense	ATGCACTTATTCTATAGTTTAA
	Antisense	TTAAACTATAGAATAAGTGCAG
Tmprss15.1100	Sense	CTACAGATGTGTTAGATATCTA
	Antisense	TAGATATCTAACACATCTGTAT
Tmprss15.12	Sense	AACGCAGCAGACTCACAGTGAA
	Antisense	TTCACTGTGAGTCTGCTGCGTG
Tmprss15.1344	Sense	ACAGGATCTTGATGATGATAAT
	Antisense	ATTATCATCATCAAGATCCTGG
Vwa5a.1345	Sense	CTAGTGTCATTAGAGAAGTTAA
	Antisense	TTAACTTCTCTAATGACACTAA
Vwa5a.1792	Sense	AAAAGTCCTTGATTCAGACTAA
	Antisense	TTAGTCTGAATCAAGGACTTTG
Vwa5a.2674	Sense	AACGGAGTTTCGAGACCATTTA
	Antisense	TAAATGGTCTCGAAACTCCGTG
Pik3cg.1153	Sense	CCACTGTGTTTGTGGAAGCGAA
	Antisense	TTCGCTTCCACAAACACAGTGA
Pik3cg.1814	Sense	ACAGCAAGAAATTGTTGCCAAA
	Antisense	TTTGGCAACAATTTCTTGCTGC
Pik3cg.3249	Sense	CAAGGATGGACTGTGCAGTTTA

	Antisense	TAAACTGCACAGTCCATCCTTT
Lama1.4983	Sense	CAAGGAGCTCACCAGAGTGTTA
	Antisense	TAACACTCTGGTGAGCTCCTTT
Lama1.5478	Sense	CAAGCTGCGTGTTCAAGAAGAA
	Antisense	TTCTTCTTGAACACGCAGCTTT
Lama1.8808	Sense	AAAGGAAGGTTACAAAGTTCGA
	Antisense	TCGAACTTTGTAACCTTCCTTG
Fzd4.1591	Sense	CCGGGACAAAGACAGACAAGTT
	Antisense	AACTTGTCTGTCTTTGTCCCGT
Fzd4.1602	Sense	CCAGACAAGTTGGAAAGGCTAA
	Antisense	TTAGCCTTTCCAACTTGTCTGT
Fzd4.1915	Sense	CAGGCAACGAGACTGTGGTATA
	Antisense	TATACCACAGTCTCGTTGCCTT
Fancc.1209	Sense	CCAGAATTGTTGATGAAATGTT
	Antisense	AACATTTCATCAACAATTCTGA
Fancc.1512	Sense	CGAGCTGGTTTTTGTTTGTTCA
	Antisense	TGAACAAACAAAAACCAGCTCT
Fancc.1921	Sense	ACAGACCTTGTACAGATCACAA
	Antisense	TTGTGATCTGTACAAGGTCTGG
Fbn2.1646	Sense	ATAGCTACAGATGCGAGTGTAA
	Antisense	TTACACTCGCATCTGTAGCTAG
Fbn2.3205	Sense	CCAGAGTGTGAAGAATGTCCTA
	Antisense	TAGGACATTCTTCACACTCTGT
Fbn2.3545	Sense	ACGGCTTCATGATGATGAAGAA
	Antisense	TTCTTCATCATCATGAAGCCGC

All the sequences listed in the above table were designed in the following cassette with miR30-backbone flanking sequences: 5′ TGCTGTTGACAGTGAGCG - Sense Strand - TAGTGAAGCCACAGATGTA - Antisense Strand - TGCCTACTGCCTCGGA 3′

2.1.8 Cell lines and bacteria

Table 8. List of the cell lines used to perform the described experiments in this study.

Cell line/Bacteria	Supplier
CaMCaA	Generated from murine HCC tissue by the group of Prof. Dr. med. Lars Zender (University of Tübingen, Germany)
HEK293T/17	ATCC, USA

MACH-1	Kind gift from Prof. Dr. Stephanie Rössler (University of Heidelberg, Germany)
NEB® Stable Competent E. coli	New England Biolabs
XL10-Gold Ultracompetent Cells	Agilent

2.1.9 Buffers and solutions

Table 9. List of the buffers and solutions and the respective recipes used in the experiments of this study.

Buffer/Solution	Component	Final concentration
Borat buffer for SDS-PAGE (pH=8.8 with NaOH)	Boric acid	20 mM
	EDTA	1.27 mM
	(in Millipore water)	
IP Sample buffer	1.0 M Tris-HCl (pH=6.8)	100 mM
	Glycerin	5% (v/v)
	10% SDS (pH=7.4)	5% (v/v)
	2-Mercaptoethanol	5% (v/v)
	Bromphenol Blue	0.005% (w/v)
	(in Millipore water)	
Laemmli Buffer with DTT (4×)	1.0 M Tris (pH=6.8)	250 mM
	Glycerol	40% (v/v)
	SDS	8% (w/v)
	DTT	100 mM
	Bromphenol Blue	0.04% (w/v)
	(in Millipore water)	
LB-Agar *	Bacto Agar	1.43% (w/v)
	(in LB-medium)	
LB-medium *	LB Broth	2.5% (w/v)
	(in Millipore water)	
Running buffer for SDS-Polyacrylamide gel electrophoresis (PAGE)	Tris	0.25 M
	Glycine	1.92 M
	20% SDS	0.5% (v/v)
	(in Millipore water)	
Stacking gel for Tris-glycine SDS-Polyacrylamide gel (5%)	ROTIPHORESE®Gel 30	17% (v/v)
	1.0 M Tris-HCL (pH=6.8)	12.5% (v/v)
	10% SDS	1% (v/v)
	10% Ammonium persulfate (APS)	1% (v/v)
	TEMED	0.1% (v/v)
	Millipore water	68%

TBS-Tween	10× TBS	1% (v/v)
	Tween 20	0.01% (v/v)
	(in Millipore water)	
Transfer buffer without Methanol for SDS-PAGE	Tris	0.25 M
	Glycine	1.92 M
	(in Millipore water)	
Tris-acetate-EDTA-Electrophoresis (TAE) Buffer (50×) (pH=8.3)	Tris	2.0 M
	Pure Acetic acid	5.71% (v/v)
	0.5 M EDTA (pH=8.0)	10% (v/v)
	(in Millipore water)	
Tris-Buffered Saline (TBS) (10×) (pH=7.6 with HCl)	Tris	200 mM
	NaCl	1.37 M
	(in Millipore water)	
Tris-glycine SDS-Polyacrylamide gel (8%)	ROTIPHORESE®Gel 30	27%
	1.5 M Tris-HCL (pH 8.8)	25%
	10% SDS	1%
	10% Ammonium persulfate (APS)	1%
	TEMED	0.06%
	Millipore water	46%
Tris-glycine SDS-Polyacrylamide gel (10%)	ROTIPHORESE®Gel 30	33%
	1.5 M Tris-HCL (pH=8.8)	25%
	10% SDS	1%
	10% Ammonium persulfate (APS)	1%
	TEMED	0.04%
	Millipore water	40%
Tris-glycine SDS-Polyacrylamide gel (15%)	ROTIPHORESE®Gel 30	50%
	1.5 M Tris-HCL (p= 8.8)	25%
	10% SDS	1%
	10% Ammonium persulfate (APS)	1%
	TEMED	0.04%
	Millipore water	23%

The recipes of buffers and solutions not listed in the above table, have been provided in the corresponding method section (refer to section 2.2). * These mediums were autoclaved before use.

2.1.10 Instruments and equipment

Table 10. List of the instruments and equipment used to perform the experiments described in this study.

Instrument/Equipment	Supplier
Agilent 2100 bioanalyzer	Agilent

Axio Vert 40c	ZEISS
Counting chamber BLAUBRAND® Neubauer	BRAND
D-DiGit® Gel Scanner	LI-COR
DISCOVERY COMFORT pipettes (10, 20, 100, 200, 1000 µL)	Corning HTL SA
Dri-Block® Heater	TECHNE
EV265 and EV231 Electrophoresis Power Supplies	Consort bvba
FluorChem™ M system	ProteinSimple
FLUOstar Omega Microplate Reader	BMG LABTECH
Gel electrophoresis chamber	FEBIKON Labortechnik
GeneChip Fluidics Station 450	Affymetrix
GeneChip Hybridization oven 640	Affymetrix
GeneChip Scanner 3000	Affymetrix
Guava easyCyte HT	EMD Millipore Corporation
Heracell™ VIOS 250i CO2 Incubator	Thermo Fisher Scientific
Heraeus Megafuge 1.0RS	Heraeus
Heraeus Megafuge 16R Centrifuge	Thermo Fisher Scientific
HI-2210 Bench Top pH Meter	Hanna Instruments
Incubator Hood TH 15	Edmund Bühler
JEM-1400Flash Electron Microscope	JEOL
KERN EW6000-1M	KERN & SOHN
Leica EM TRIM 2.	Leica
Magnetic stirrer C-MAG MS 7	IKA®-Werke
Memmert 37°C Bacteria Incubator	Memmert
Fluorescence microscope BX53	OLYMPUS
MIKRO 200 R	Andreas Hettich
Mini Trans-Blot® Cell	BioRad Laboratories
Mini-PROTEAN® Tetra Cell	BioRad Laboratories
Mr. Frosty™ Freezing Container	Thermo Fisher Scientific
myFUGE™ Mini centrifuge	Benchmark Scientific
NanoDrop™ ND-1000 UV/Vis Spectrophotometer	NanoDrop Technologies
Odyssey® DLx Imaging System	LI-COR Biosciences
Orbitrap Exploris 480 Mass Spectrometer	Thermo Fisher Scientific
PIPETBOY acu 2	INTEGRA Biosciences
PowerPac™ Basic Power Supply	BioRad Laboratories

ProteinSimple FluorChem E Imaging	Alpha Innotech Corp.
PTC 200 Peltier Thermal Cycler	MJ Research
QuantStudio 3 Real-Time PCR System	Thermo Fisher Scientific
Revolver Rotator, digital (D-6050)	neoLab
Roller mixer RN 5	Ingenieurbüro CAT, M. Zipperer
Seahorse XFe96 Analyzer	Agilent Technologies
SW22 Shaking water bath	JULABO USA
Thermomixer Compact 5350 Mixer	Eppendorf
Transsonic T460/H ultrasonic bath	Elma Schmidbauer
UNITWIST-RT benchtop agitator	Kisker Biotech
VACUSAFE vacuum pump	INTEGRA Biosciences
Vortex Mixer 7-2020	neoLab

2.1.11 Computer software

Table 11. List of the computer software used in performing the experiments described in the results section and analyzing the associated data.

Software	Supplier
ANNOVAR	http://www.openbioinformatics.org/annovar/
cellSens Dimension	OLYMPUS
DNADragon 1.6.0.	SequentiX - Digital DNA Processing
GraphPad Prism8	GraphPad Software Inc.
Guava® InCyte™ Software	Luminex
Image Studio Lite Version 5.2.5	LI-COR
ImageJ (Fiji)	National Institutes of Health, USA
Ingenuity® Pathway Analysis (IPA®)	QIAGEN
Inkscape	https://inkscape.org/ (The Inkscape Project)
JMP11 Genomics Version 7	Statistical Analysis System (SAS)
MARS Data Analysis Software	BMG LABTECH
Microsoft Excel	Microsoft
NanoDrop 1000 3.8.1	Thermo Fisher Scientific
R Studio	RStudio, PBC.
Seahorse Wave Desktop Software (V2.6.1.53)	Agilent Technologies
QuantStudio™ Design & Analysis Software v1.4.3	Thermo Fisher Scientific

2.2   Methods

2.2.1   Human material

A formalin-fixed paraffin-embedded (FFPE) human HCC specimen showing a nodule-in-nodule pattern of a well-differentiated HCC with surrounding premalignant HGDN was selected for whole exome sequencing. Normal gallbladder tissue of the same patient was used to remove germline variants. The study was approved by the local ethics committee of the Medical Faculty of the University of Heidelberg (S-462/12).

2.2.2   Human tissue DNA extraction

Hepatocellular carcinoma (HCC), Dysplastic Nodule (DN), cirrhotic liver (CL), and gallbladder (GB) tissues were manually microdissected from six consecutive 10-μm thick FFPE sections. For DNA extraction, microdissected samples from each tissue were pooled, deparaffinized, and digested with proteinase K overnight. The extraction was performed with an automated Maxwell 16 Research extraction system (Promega) using the Maxwell$^{®}$ 16 FFPE Plus LEV DNA Purification Kit following the manufacturer's instructions. The concentration of DNA was measured fluorometrically using the Qubit$^{™}$ 2.0 DNA high sensitivity kit (Thermo Fisher Scientific).

2.2.3   Whole Exome Sequencing

Whole Exome Sequencing (WES) of the samples was done by Dr. Robert Geffers (Genome Analytics, Helmholtz Centre for Infection Research Braunschweig). For preparation of the sequencing library, the concentration and quality of the extracted genomic DNA (gDNA) was determined by a 2100 Bioanalyzer (Agilent Technologies). Then, to obtain DNA fragments with an average length of 300 base pairs (bp), 200 ng of the gDNA (diluted in 55μl Tris-EDTA buffer) was fragmented in a Covaris microTUBE using a Covaris S2 ultrasonicator (Covaris, Inc.) with the operating condition of duty cycle 10%, intensity 4, and 200 cycles per burst during 80 seconds. The size of the resulting fragments was controlled by the 2100 Bioanalyzer (Agilent Technologies).

The DNA sequencing library was generated from the fragmented gDNA (200 ng) using the SureSelect Human All Exon v.2 (44Mb) kit (Agilent Technologies), according to the instructions of the manufacturer. The library was then purified and the size was controlled using the Agilent Technologies 2100 Bioanalyzer (High Sensitivity DNA Chip). Next, it was prepared for sequencing following the protocol provided by the manufacturer (Illumina). The

sequencing was performed on an Illumina Genome Analyzer IIx system using 2× 75 bp paired end run with an average of 12.5×10^6 reads per single exome (mean coverage of 20-50×).

Quality of FASTQ files was controlled by FastQC tool. Raw reads were further trimmed by fastq-mcf based on default settings, minimum length of 50 bp, and Phred quality score > 20, to eliminate artificial Illumina adapter sequences as well as bad quality sequence reads. Trimmed FASTQ files were aligned to human reference genome hg19 using BWA aligner tool (http://bio-bwa.sourceforge.net/) followed by tagging duplicated reads (PCR products) with MarkDuplicates (Picard tools v1.55). Resulting BAM files were processed via GenomeAnalysisTK-1.4 (https://gatk.broadinstitute.org/hc/en-us). Determination of Single Nucleotide Polymorphism (SNP) and Insertion–deletion mutations (InDels) were performed by UnifiedGenotyper (GenomeAnalysisTK-1.4). SNV annotation was conducted by applying the ANNOVAR pipeline (version 2012-03-08) [116] using the following databases: dbSNP132, EPS5400_ALL and 1000GenomesProject (1000g2010nov_ALL).

Single Nucleotide Polymorphism (SNP) and Insertion–deletion mutations (InDels) were normalized against the normal gall bladder control. Then, common polymorphisms listed in dbSNP132 and variants of the projects 1000 genomes and EPS5400 were eliminated. Next, the variants were selected for being non-synonymous, exonic (somatic or loss of heterozygosity (LOH) with p-value <0.05), and functionally damaging (PolyPhen-2 ≥0.85). Reads were checked visually and identified clonally expanded variants were validated by Sanger sequencing. Integrative Genomics Viewer was used for visualization. [117] Data analyses of this part were all performed by Dr. Robert Geffers and his team.

2.2.4 Transposon vector and shRNA cloning

Design and cloning of the miR30-based shRNA library (Table 7) and the transposon vectors used in *in vivo* experiments of the study (refer to 3.2 and 3.3) were done by Dr. Torsten Wuestefeld in the research group of Prof. Dr. med. Lars Zender (University of Tübingen). 51 shRNAs (3 per gene) were transferred from the original pSM2 vector into a murine stem cell virus (MSCV)-SV40-*GFP* vector. The shRNAs were then subcloned into the transposon vector p/T-Caggs-*Myc*-IRES-*AKT1*[Myr]-IRES-*GFP*-mir30-5′ with a 1000-fold representation of each shRNA following a pooled cloning approach. [84]

2.2.5 Hydrodynamic tail vein injection

Hydrodynamic tail vein injection of the shRNA library (refer to 2.2.4 and 3.2) or the individual variants of the potential driver genes of hepatocarcinogenesis (refer to 3.3) into *Trp53* heterozygous mice with the C57BL/6 background was conducted in collaboration with Prof. Dr. med. Lars Zender and his research group (University of Tübingen). Briefly, 25 µg DNA dissolved in a saline solution equivalent to 10% of the mouse body weight was injected via the tail vein in < 10 seconds. Plasmids were prepared using the QIAGEN EndoFree Maxi Kit (QIAGEN). Transposon and sleeping beauty transposase-encoding plasmids were used in a molar ratio of 5:1 respectively. [84] All the animal experiments in this study were conducted following the Animal Welfare Guidelines of Germany and upon approval of the German legal authorities.

2.2.6 Monitoring and tumor sample collection in mice

Hydrodynamically tail vein injected mice were monitored for tumor progression by abdominal palpation and entire-body GFP visualization in collaboration with Prof. Dr. med. Lars Zender and his research group (University of Tübingen). Following the tumor progression, mice were sacrificed to harvest HCC nodules. In shRNA library screening experiment (see 3.2), genomic DNA was isolated from the harvested nodules to be used for NGS and identification of enriched shRNAs. In single variant validation experiment (see 3.3), HCC nodules were in part fixed in formalin and then histological sections were stained with hematoxylin and eosin (H&E) following paraffin-embedding. Histopathological evaluations and typing of the nodules were conducted by Prof. Dr. med. Thomas Longerich (University Hospital Heidelberg).

2.2.7 Next Generation Sequencing

The shRNAs enriched in the HCC nodules were identified in collaboration with Prof. Dr. med. Lars Zender and his research group (University of Tübingen). For this, genomic DNA was isolated from the nodules. The Illumina adaptor sequence (flanking the miR30 cassette) was used to amplify the transposon sequences integrated into the genome. Single-end sequencing run with a read length of 46 bp was performed using the Illumina Genome Analyzer IIx system in the HZI array and genome analysis facility (Braunschweig, Germany). The Illumina Pipeline Version 1.8 was used for base calling and export of the multi FASTA format sequencing data. Data analysis was conducted using a Perl-Script (M.J.) and the retrieved sequences were matched with the shRNA library data. [84]

2.2.8 Cloning

The murine full length *Prss23* wild-type cDNA (clone BC018517) inserted in pCMV-SPORT6 vector was purchased from Open Biosystems (#MMM1013-65843). The *Prss23* insert was then transferred from the pCMV-SPORT6 vector into the pCaggs vector (kindly provided by Prof. Dr. med. Lars Zender). First, 6µg of pCMV-SPORT6-*Prss23* was digested with NotI and SalI restriction enzymes and then loaded into 1% agarose gel. The separated *Prss23* cDNA was cut out of the gel and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega), following the manufacturer´s instructions. In parallel, 2 µg of the pCaggs plasmid was digested with the same restriction enzymes used for digestion of *Prss23* cDNA and the same procedure for gel purification was applied. To subclone *Prss23* wild-type into pCaggs plasmid, ligation reaction was performed using the T4 DNA Ligase (NEB) in a molar ratio 1:3 of vector to insert. For this reaction, 100 ng pCaggs vector and 103.7 ng insert (*Prss23*) were mixed and 1 µl of T4 DNA Ligase and 4 µl of T4 DNA Ligase Buffer (10×) were added to the mix filled up with nuclease-free water up to total volume of 30 µl. The reaction mix was incubated overnight at 16°C in a thermal cycler. 2 µl of the ligation mixture was used to transform 50 µl NEB® Stable Competent *E. coli* (C3040) (NEB) following the competent bacteria transformation protocol (for details refer to 2.2.10). Selection of successfully transformed bacteria was done using Ampicillin. Bacteria colonies were picked on the following day and were cultured in 3 ml Lysogeny broth (LB)-Medium with antibiotic (Ampicillin or Carbenicillin 100 µg/ml) in sterile Falcon™ Round-Bottom tube overnight at 37 °C in a shaking incubator (250 rpm). Plasmid DNA was isolated from the bacterial culture using Monarch® Plasmid Miniprep Kit (NEB) following the instructions of manufacturer. The length of extracted plasmid was validated using agarose gel electrophoresis. Next, the sequence of subcloned *Prss23* wild-type in pCaggs vector was validated by Sanger sequencing using the sequencing primers indicated in Table 6.

QuikChange® Lightning Site-Directed Mutagenesis Kit (Agilent) was used to introduce the identified P230A mutation (murine *Prss23* BC:c.686C>G) in sequence of *Prss23* wild-type in pCMV-SPORT6 vector, according to the manufacturer's instructions. Primers were designed using the QuikChange® Primer Design Program (Agilent) [118] and their sequence is listed in Table 6. The introduction of single nucleotide mutation in the sequence of *Prss23* was validated by Sanger sequencing using the sequencing primers detailed above. Mutated *Prss23* was subcloned into pCaggs backbone as described above for the wild-type *Prss23* variant.

In order to fuse a FLAG tag sequence, each of the *Prss23* wild-type and the *Prss23*[P230A] variant were separately subcloned into a pDEST26-FLAG-C-Puro vector (kindly provided by Prof. Dr. Stephanie Rössler) using the Gateway® Technology (Invitrogen). For detailed description refer to 2.2.9 Gateway® cloning. NEBuilder® HiFi DNA Assembly Master Mix was used to subclone *Prss23* C-terminal FLAG-tagged variants into the lentiviral vector pLV-EF1a-IRES-Blast. C-terminally FLAG-tagged *Prss23* variants were PCR amplified from the respective pDEST26-*Prss23*[WT] or *Prss23*[P230A]-FLAG-C-Puro vector using forward and reverse primers introducing BamHI-HF and EcoRI-HF restriction sites to the N- and C-terminal sites of the *Prss23* sequence, respectively. The primers were designed using the NEBuilder® Assembly Tool (v2.2.8) and a minimum primer length of 18 nucleotides and a minimum overlap of 20 nucleotides was set for primer design (https://nebuilder.neb.com/). The destination vector pLV-EF1a-IRES-Blast (1µg/reaction) was digested with restriction enzymes BamHI-HF® (1µl/reaction) and EcoRI-HF® (1µl/reaction) in CutSmart buffer (up to 50 µl/reaction) at 37°C for 1 hour. pLV-EF1a-IRES-Blast was a gift from Tobias Meyer. (Addgene plasmid #85133; http://n2t.net/addgene:85133; RRID: Addgene_85133) [119]

The restriction enzyme (RE) digested plasmid (100 ng/reaction) and gel purified PCR product (vector-fragment molar ratio of 1:2) were mixed with NEBuilder® HiFi DNA Assembly Master Mix (10 µl/reaction) and deionized H_2O (up to 20 µl/reaction) in a 1.5 ml microcentrifuge tube and incubated at 50°C for 15 minutes. 2 µl of the chilled assembled product were used to transform 50 µl NEB® Stable *Mix & Go* Competent *E. coli* (C3040). The transformation reaction was performed as detailed in 2.2.10 Bacteria transformation. Ampicillin was used for antibiotic selection of successfully transformed bacteria. The following day, bacteria colonies were picked and cultured for plasmid DNA isolation as described above.

In all cloning experiments of this study, 1 ml of overnight bacterial cultures was added to 330 µl autoclaved Glycerin solution in water (60%) in 2 ml cryotubes and stored at -80 °C. Plasmid DNA stocks were prepared from sequence-verified plasmids. For this, plasmid was isolated from 350 ml overnight bacterial culture using the PureYield™ Plasmid Midiprep System (Promega) following the instructions from the manufacturer.

2.2.9 Gateway cloning

The Gateway® Technology (Invitrogen) was employed to subclone *Prss23* wild-type and P230A variants into pDEST26-FLAG-C-Puro (kindly provided by Prof. Dr. Stephanie

Rössler). For this, the sequences of *Prss23*^{WT} and *Prss23*^{P230A} were PCR amplified from pCaggs-*Prss23*^{WT} or -*Prss23*^{P230A} vectors using Gateway® attB1 and attB2 primers and removing the stop codon from the gene sequence. The primers were designed following the instructions from the manufacturer and are listed in Table 6. The PCR products were validated and purified as described earlier (refer to 2.2.8).

First, the BP Clonase™ reaction was performed to subclone *Prss23*^{WT} and *Prss23*^{P230A} into the donor vector pDONR201 (kindly provided by Prof. Dr. Stephanie Rössler). For this, purified PCR products (40 ng) were separately mixed with pDONR201 (100 ng) and TE-Buffer (pH=8.0, up to 8 µl) in a 1.5 ml tube at room temperature (RT). 2 µl of the Gateway™ BP Clonase™ II Enzyme Mix was added to each reaction and the tubes were briefly vortexed twice and later incubated at 25°C for 6 hours. Afterwards, the reactions were terminated by adding 1 µl proteinase K solution to each tube followed by 10 minutes incubation at 37°C. 5 µl of the reaction was used to transform 50 µl *Mix & Go* competent MACH-1 bacteria. The mixture of plasmid and bacteria was incubated on ice for 60 minutes. Then, 180 µl LB-Medium was added to the mixture and incubated at 37°C with shaking (350 rpm) for 60 minutes. 20 µl of the transformed bacteria culture was plated on a pre-warmed antibiotic-coated (Kanamycin 25 µg/ml) LB-agar plate using sterile glass beads. LB-agar plates were incubated overnight at 37 °C. The following day, bacteria colonies were picked and cultured for plasmid DNA isolation as described before (refer to 2.2.8).

Second, the LR Clonase™ reaction was performed to transfer the *Prss23*^{WT} and *Prss23*^{P230A} from the entry clone (pDONR201-*Prss23*$^{WT/P230A}$) into the destination vector (pDEST26-FLAG-C-Puro) and to generate an expression clone (pDEST26- *Prss23*^{WT} or *Prss23*^{P230A}-FLAG-C-Puro). For this reaction, 1 µl pDONR201-*Prss23*^{WT} or *Prss23*^{P230A} (75 ng/µl) and 1µl pDEST26-FLAG-C-Puro (30 ng/µl) were mixed in 2 µl TE-Buffer (pH=8.0) in a 1.5 ml microcentrifuge tube at RT. 1 µl of Gateway™ LR Clonase™ II Enzyme Mix was added to each reaction. The tubes were shortly vortexed two times and were incubated at 25°C for 6 hours. 0.5 µl proteinase K solution was added to each tube followed by 10 minutes incubation at 37°C to terminate the reaction. The LR Clonase™ reaction (5.5 µl) was used to transform 50 µl *Mix & Go* competent MACH-1 bacteria. The transformation of bacteria was performed as described for the BP Clonase™ reaction with the exception of using Ampicillin (100 µg/ml) for antibiotic selection. The following day, bacteria colonies were picked and cultured for plasmid DNA isolation detailed earlier (refer to 2.2.8).

2.2.10 Bacteria transformation

To transform bacteria, plasmid DNA (100 ng) was added to 50 µl chemically competent bacteria and was mixed by flicking the tube 4-5 times followed by incubation on ice for 2-10 minutes. Then, the mixture was heat shocked at 42°C for exactly 30 seconds and incubated on ice for 2 minutes. Next, room temperature S.O.C outgrowth medium (up to 1 ml) was added to the mixture and incubated at 37°C with shaking (250 rpm) for 60 minutes. Bacterial suspension (20-100 µl) was spread over the surface of pre-warmed antibiotic coated (Ampicillin or Carbenicillin 100 µg/ml and Kanamycin 25 µg/ml) LB-agar plate using sterile glass beads or a sterile spreader. The LB-agar plate was incubated overnight at 37 °C (not longer than 16 hours).

2.2.11 Generation of *Mix & Go* bacteria

In order to generate chemically competent *E.coli,* the *Mix & Go* transformation kit from Zymo Research was used. For this, bacteria were cultured overnight in LB-medium. The following day, ZymoBroth™ growth medium was added to the bacterial culture (in 1:10 ratio) followed by shaking (250 rpm) at RT. The absorbance of the bacterial culture was constantly measured at 600 nm until it reached 0.4. Then, the culture was incubated on ice for 10 minutes and centrifuged at 3000 rpm for 10 minutes at 4°C. The supernatant was discarded and the bacteria pellet was resuspended in 5 ml ice-cold 1× Wash Buffer (diluted in Dilution Buffer) provided in the kit. The cells were re-pelleted as before and gently resuspended in 5 ml ice-cold 1× Competent Buffer (diluted in Dilution Buffer). The competent bacteria were aliquoted and stored at -80°C.

2.2.12 Lentivirus production

HEK293T cells were seeded (10^6 cells per T25 flask) 24 hours prior to plasmid transfection. On the following day, 3.33 µg of psPAX2 and 1.04 µg of pMD2.G and 4.17 µg of lentiviral pLV-EF1a-IRES-Blast plasmids, expressing either the wild-type or P230A variant of *Prss23* with/without C-terminal FLAG tag, were added to 417 µl OptiMEM. Latest, 25 µg Polyethylenimine (PEI) was added to the plasmid mix and vortexed for 5 seconds followed by 30 minutes incubation at RT. Meanwhile, the old medium of HEK293T cells was aspirated and replaced with 6.5 ml of fresh CaMCaA complete medium after washing the cells with 10 ml PBS. The transfection mix was added to the cells in a dropwise manner and the flasks were incubated for 24 hours. On the following day, medium was replaced with fresh pre-warmed medium. Medium containing viral particles was collected 48 hours post transfection. The

supernatant was filtered using sterile syringe filters with cellulose acetate membrane (pore size 0.45 µm) and aliquoted in 1 ml and frozen at -80 °C for long-term storage. psPAX2 and pMD2.G were a gift from Didier Trono (Addgene plasmids #12260 and #12259 respectively; http://n2t.net/addgene:12260; and http://n2t.net/addgene:12259; RRID: Addgene_12259 and Addgene_12260).

2.2.13 Establishment of isogenic cell lines

The primary murine HCC cell line CaMCaA was established from a murine HCC generated by expression of a transposon vector expressing murine MYC and human myristoylated AKT1 (pCaMIA) in a *Trp53* heterozygous mouse (C57BL/6 background). It was provided by Prof. Dr. med. Lars Zender. CaMCaA parental cells were seeded in 6-well plate (40 000 cells per well) 24 hours prior to lentiviral transduction. On the following day, cells were 50% confluent. Medium of cells was aspirated and 2 ml pre-warmed fresh medium was added to each well. CaMCaA cells were treated with polybrene (5 µg/ml) to improve transduction efficacy and were infected with 1 ml lentiviral supernatant. CaMCaA cells were transduced with lentiviral particles carrying either the plasmid overexpressing mouse *Prss23* wild-type or mouse *Prss23* P230A variant fused with/without C-terminal FLAG Tag. Cells were incubated at 37°C for 72 hours. Thereafter, medium was aspirated from wells and cells were washed 3 times with PBS and pre-warmed fresh medium was added to each well. Cells were expanded from 6-well plate into 10-cm plate on the following day and were treated with Blasticidin S HCL (0.4 µg/ml) to select successfully transduced cells. Expression of PRSS23 variants was validated at mRNA and protein levels (refer to 2.2.20 and 2.2.21). Following that, single cell clones were cultured in order to establish isogenic cell lines. For this aim, highly diluted cell suspensions (1 cell/ 100 µl) in Blasticidin supplemented medium (1.5-1.8 µg/ml) were prepared and seeded in 96-well plates (100 µl per well). The growth of single clones was monitored continuously and only the wells containing a single clone were expanded into 24-well plate. When transferred to 6-well plate, single cell clones were validated for overexpression of PRSS23 at mRNA and protein levels as described before.

2.2.14 Cell culture

The cell lines listed in Table 8 were cultured in Gibco™ DMEM supplemented with 10% (v/v) Gibco™ FBS and 1% (v/v) penicillin/streptomycin and 1% (v/v) Gibco™ MEM NEAA (100×) and 1% (v/v) Gibco™ Sodium Pyruvate (100 mM) (termed as CaMCaA complete medium later in this thesis) at 37 °C with 5% CO_2. At 90-100% confluency cells were

passaged. Medium was aspirated and cells were washed with PBS. Then cells were incubated with Gibco™ Trypsin-EDTA (0.05%) for 3-5 minutes depending on the cell line at 37 °C in cell culture incubator. The incubation time required for detachment was ~5 minutes for CaMCaA cells and ~3 minutes for HEK 293T cell. Detached cells were collected in pre-warmed CaMCaA complete medium. Cell suspension containing a known number of cells depending on the size of culture dish and the growth rate of the cell line were added to fresh CaMCaA complete medium. All cell lines were routinely tested for mycoplasma contamination using the MycoAlert™ Mycoplasma Detection Kit (Lonza).

2.2.15 Freezing and thawing of cell lines

For freezing, cells cultured in 10-cm plates were washed with 5 ml PBS and were trypsinized with 1 ml Gibco™ Trypsin-EDTA (0.05%) at 90-100% confluency. Cells were resuspended in 8 ml CaMCaA complete medium and centrifuged at 1000 rpm for 5 minutes. Medium was aspirated and cell pellet was resuspended in 1620 µl CaMCaA complete medium and were transferred into 2 ml cryotubes. 180 µl DMSO (1:10) was added to the cell suspension and the tube was stored in freezing container at -80 °C for 24 hours. The following day, cells were transferred into liquid nitrogen (-196 °C) for long-term storage.

To thaw up, frozen cells were quickly thawed up by incubation in water bath at 37 °C for 30-60 seconds. The 1.8 ml cell suspension was dissolved in 9 ml prewarmed medium and centrifuged at 1000 rpm for 5 minutes. Medium was aspirated and cell pellet was resuspended in 10 ml pre-warmed fresh CaMCaA complete medium and cultured in 10-cm plate.

2.2.16 Sanger sequencing

Sequencing of samples was done by Microsynth Seqlab (Switzerland). 480-1200 ng of the plasmid DNA or 18 ng/100 bp of the PCR product diluted in 12 µl ultrapure DNase/RNase free distilled water mixed with 3 µl sequencing primer (10 µM) was sent for sequencing. The DNA samples were gel or PCR purified samples. The sequencing results were assembled together with the reference sequence using the software DNADragon v1.6.0 (www.dna-dragon.com) and were visually verified by reference sequence comparison.

2.2.17 Cell viability assay

CaMCaA cells expressing either PRSS23^{WT} or PRSS23^{P230A} were cultured in 96-well plates (2000 cells per 100 µl medium per well) for 24, 48, and 72 hours. At each timepoint, the medium was aspirated and 100 µl of Methylthiazolyldiphenyl-tetrazolium bromide (MTT)

solution (1:10 dilution in CaMCaA complete medium) was added to each well and incubated at 37°C in an incubator for 1 hour. The medium was aspirated from the wells and 100 µl DMSO:EtOH solution (1:1) was added to each well to dissolve intracellular MTT formazan salts. Subsequently, the plate was shaken for 5 minutes before the absorbance was measured at 750 nm using the FLUOstar® Omega Microplate Reader (BMG LABTECH).

2.2.18 Invasion assay

Transwell invasion assays were performed using the Corning® BioCoat™ Growth Factor Reduced Corning® Matrigel® Invasion Chamber (Corning) to study the invasion capacity of PRSS23^{WT} or PRSS23^{P230A} expressing CaMCaA cells. Two hours prior to the assay, the medium of cells (at confluent state) was replaced with FBS-free medium (CaMCaA complete medium without FBS) and the inlay chambers were equilibrated in FBS-free medium. For this, 750 µl FBS-free medium was added to each well of a 24-well plate. Using a sterile forceps, inlays were carefully placed into the wells without trapping bubbles between the inlay and the medium surface. 100 µl FBS-free medium was pipetted into the inlay and the 24-well plate was incubated in cell culture incubator for at least 2 hours.

Following that, cells were trypsinized and collected in 8 ml FBS-free medium. The cell suspension was centrifuged at 1000 rpm for 5 minutes. Then, the medium was aspirated and the pellet was dissolved in 8 ml FBS-free medium. Cells were counted and a cell suspension (60 000 cells per ml) in FBS-free medium was prepared. Meanwhile, the medium inside equilibrated inlays was aspirated and the inlays were transferred into another 24-well plate containing fresh 750 µl CaMCaA complete medium (with FBS 10% (v/v)). 500 µl of the cell suspension was seeded dropwise in each inlay and incubated at 37°C in the CO$_2$ cell incubator for 24 hours.

Medium was aspirated from inlays and the inside of inlays was washed with PBS and gently swabbed with cotton swabs to remove non-invaded cells. To fix the invaded cells on the bottom side of the membrane, inlays were placed in a 24-well plate containing 750 µl 4% PFA (v/v in PBS) for 15 minutes at RT. Next, inlays were placed in another 24-well plate containing 750 µl 0.1% crystal violet stain solution (w/v in 20% MeOH) for 10 minutes at RT. After that, inlays were rinsed well 3 times with PBS and air dried before visualization using an OLYMPUS BX53 microscope (OLYMPUS).

Densitometric quantifications of microscopic images (10× magnification) of crystal violet-stained membranes were done by Fiji (NIH). [120] At least five randomly selected fields

on each membrane were analyzed for the total area covered by crystal violet-stained cells. Two technical replicates per clone and three independent biological replicates were considered for this experiment.

2.2.19 Real-Time ATP Rate Assay

To study the bioenergetic profile of CaMCaA cells expressing PRSS23^{WT} or PRSS23^{P230A}, the Seahorse XF Real-Time ATP Rate Assay (Agilent) was performed following the instructions from the manufacturer. CaMCaA cells (1500 cells per 100 µl per well) were homogenously seeded in Poly-D-Lysin solution (22.4 µg/ml) coated seahorse microplate 72 hours prior to the assay and were cultured in normal cell culture incubator (37°C and 5% CO_2) until the day of the assay. Treatment of cells with Dimethyloxallyl Glycine (DMOG) was done for one hour prior to the assay during the incubation in a non-CO_2 incubator. DMOG was dissolved in assay medium (final concentration 1 mM). After the assay, assay medium was discarded and cells were gently washed with ice-cold PBS (1-2 times). 20 µl cell lysis buffer (cell lysis buffer supplemented with 1% PMSF, 1% protease- and 10% phosphatase-inhibitors) was added to each well and 3 µl cell lysate (in duplicates) was used for Bradford protein quantification assay. The protein concentrations were used for normalization using the Seahorse Wave Desktop Software 2.6.1 (Agilent Technologies).

2.2.20 RNA extraction, cDNA synthesis, and quantitative real-time PCR

Cultured cells were washed with ice-cold PBS and were harvested from 6-well plates in 1.5 ml Eppendorf tubes with Lysis Buffer RA1 (MACHEREY-NAGEL) supplemented by 1% β-mercaptoethanol (350 µl/well) using a pre-chilled plastic cell scraper. RNA was extracted from the total cell lysates using the NucleoSpin RNA kit (MACHEREY-NAGEL), according to the manufacturer's instruction. The concentration of RNA samples was determined by NanoDrop™ ND-1000 UV/Vis Spectrophotometer (NanoDrop Technologies).

cDNA synthesis was done by reverse transcription (RT) of the total RNA. 1 µg RNA (a total volume of 12 µl mixed with water/reaction) was added to Random primer (0.5 µM) and dNTPs (1 mM) in a PCR tube on ice and incubated at 75°C for 5 minutes in a thermocycler. 4 µl of 5× Reaction Buffer for RT (Thermo Scientific) was added to each tube on ice and incubated at 25°C for 5 minutes in a thermocycler. Then, 1 µl of Thermo Scientific™ RevertAid H Minus Reverse Transcriptase (200 U/µl) was added to each tube on ice and incubated at 25°C for 10 minutes followed by incubations at 42°C for 1 hour and at 70°C for 10 minutes in the thermal cycler.

Quantitative real-time polymerase chain reaction (qRT-PCR) was performed to determine the relative mRNA expression of genes of interest using a QuantStudio™ 3 Real-Time PCR System (Thermo Fischer Scientific). The amplification reaction mix (8 µl/reaction) was prepared using 5 µl PowerUp™ SYBR™ Green Master Mix (Thermo Fischer Scientific), 0.4 µl of each sequence-specific forward and reverse primers (10 µM) listed in Table 6, and 2.2 µl nuclease-free H_2O. 2 µl cDNA template (4 ng) was added to 8 µl reaction master mix in each well. The amplification program consisted of a pre-heating at 50°C for 2 minutes and an initial denaturation step at 95°C for 10 minutes followed by 40 cycles of denaturation (at 95°C for 15 seconds) and annealing (at 60°C for one minute). Fluorescence was recorded during each annealing step. At last, a melting step was performed which included 15 seconds of 95°C, one minute of 60°C, and 15 seconds of slow heating up to 95°C (at the rate of 0.15°C per second) during which the fluorescence intensity was continuously measured. Samples belonging to the same comparison group were run together within one qPCR experiment to avoid inter-run variations. The delta-delta Ct ($2^{-\Delta\Delta CT}$) method was used to calculate the relative gene expression of samples. *Gapdh* was used as housekeeping gene in all qPCR analysis performed in this study.

2.2.21 Protein extraction and Western immunoblotting

Medium was aspirated from cultured cells and ice-cold PBS was used to wash the cells. Cell were harvested with ice-cold cell lysis buffer (70 µl for 6-well plate and 120 µl for 10-cm dish at 90-100% confluency) supplemented with 1% (v/v) PMSF, 1% (v/v) protease- and 10% (v/v) phosphatase inhibitors. Pre-chilled plastic cell scraper was used to collect the cells pre-treated with the cell lysis buffer for 2 minutes on ice. The total cell lysates were collected into pre-chilled 1.5 ml Eppendorf tubes and were ultrasonicated for three times in a water bath (each time for 30 seconds). The tubes were incubated on ice for one minute followed by a short vortex between every sonication step. The cell lysates were then centrifuged at maximum speed for 10 minutes at 4 °C. The supernatant containing the total protein was transferred into a fresh pre-chilled tube and stored at -20°C for short-term and -80°C for long-term storage. Subcellular protein fractionation was done using the Mitochondria Isolation Kit for Cultured Cells (Thermo Fisher Scientific) for mitochondrial protein fraction and the NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific) for nuclear and cytoplasmic protein fraction, following the instructions from the manufacturer. The protein concentration was measured using NanoDrop™ ND-1000 UV/Vis Spectrophotometer (NanoDrop Technologies).

The protein samples were prepared for sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) using 4× Laemmli sample buffer. Desired amount of protein diluted in water was mixed with 8 µl of 4× Laemmli sample buffer to reach a total volume of 32µl. Diluted protein samples were boiled at 95°C for 5 minutes. Based on the size of target protein, either 8%, 10%, or 15% SDS-polyacrylamide gels were used to separate proteins. SDS-PAGE was performed at 75-90V until the Bromophenol Blue dye was running out of the gel. Next, the Western blot transfer sandwich was assembled using nitrocellulose membrane and the transferring step was done either with Borat buffer or Transfer buffer supplemented with 20 % Methanol (v/v) at 90V for 1 hour and 20 minutes. The membrane was rinsed briefly in Millipore water and was blocked with 5% (w/v) Bovine Serum Albumin (BSA) or milk in 1×TBST at RT for 1 hour. The membrane was then incubated with primary antibody dilution in blocking buffer (at recommended dilution by manufacturer) at 4°C overnight. On the following day, the membrane was rinsed 3 times (each 10 minutes) in 1×TBST at RT and incubated with fluorescent-dye conjugated secondary antibody dilution in blocking buffer (1:20 000) at RT for 1 hour. The blot was rinsed 3 times (each 10 minutes) in 1×TBST at RT before visualization. The membrane was scanned for fluorescent signals at 700 nm (red) and 800 nm (green) channels using the Odyssey® DLx Imaging System (LI-COR). All incubation steps were done with gentle rotation and the membrane was protected from light during the incubation with secondary antibody until visualization. The intensity of signals was quantified using the Image Studio Lite Ver 5.2 software (LI-COR) and was normalized to signal intensity of ACTIN or GAPDH as the loading control.

2.2.22 Polymerase chain reaction (PCR)

Polymerase Chain Reaction (PCR) using Phusion® High-Fidelity DNA Polymerase (NEB) was done for amplification of desired DNA sequences. The PCR reaction mix containing DNA template (50-250 ng genomic DNA and 1 pg-10 ng plasmid or viral DNA), dNTPs (200 µM), target sequence specific forward and reverse primers (each 0.5 µM), and Phusion DNA Polymerase (1.0 unit/50 µl reaction) in 1× Phusion HF or GC Buffer was prepared on ice. The Phusion DNA polymerase was added at last to impede primer degradation due to the $3' \rightarrow 5'$ exonuclease activity as recommended by the manufacturer. The PCR tubes were placed in the pre-heated block (98°C) of a thermocycler. Immediately after, the thermocycling consisted of the following steps was started: 30 seconds of initial denaturation at 98°C, 30 cycles of three denaturation (10 seconds at 98°C), annealing (30 seconds at primer Tm), and extension (15-30 seconds per kb at 72°C) steps, followed by 5 minutes of final

extension at 72°C. The PCR products were stored at 4°C for short-term and at -20°C for long-term storage. Primer Melting Temperature (Tm) or annealing temperature was calculated exclusively for each reaction using the NEB Tm Calculator (New England Biolabs; https://tmcalculator.neb.com/).

2.2.23 Enzyme-linked immunosorbent assay (ELISA)

Enzyme-linked immunosorbent assay (ELISA) was performed to detect and quantify PRSS23 protein in the culture medium of PRSS23^{WT} and PRSS23^{P230A} expressing CaMCaA clones. CaMCaA cells expressing PRSS23^{WT} and PRSS23^{P230A} (500 000 cells per 10-cm dish) were cultured for 72 hours. Then the culture medium was collected and centrifuged at 1000 rpm for 20 minutes. The medium (8 ml per clone) was concentrated at 4000 ×g for 20 minutes using Amicon$^{®}$ Ultra-15 (10 kDa cutoff) centrifugal filters (Merck Millipore). The concentration factor was calculated for each sample. The concentrated mediums were used for quantitative measurement of PRSS23 using the Mouse Serine Protease 23 (PRSS23) ELISA Kit (MyBioSource) following the instructions from the supplier. The absorbance was measured at 450 nm using the FLUOstar$^{®}$ Omega Microplate Reader (BMG LABTECH). Four independent experiments, each with three technical replicates, were evaluated for each CaMCaA clone. Four Parameter Logistic (4PL) curve fit model was used for quantitative analysis of data. The standard curve was plotted using the 4-parameter fit algorithm and the concentration of samples was determined according to the standard curve using the Four Parameter Logistic Curve online data analysis tool (MyAssays Ltd.). [121] The calculated concentrations were normalized to concentration factor of respective sample.

2.2.24 Immunoprecipitation

CaMCaA cells expressing C-terminally FLAG-tagged PRSS23^{WT} and PRSS23^{P230A} were cultured (350 000 cells per 10-cm plate) for 72 hours. Cells were washed with ice-cold PBS and were harvested with 200 µl of an ice-cold lysis buffer containing 1 ml Tris-HCL (pH=6.8, 1.5 M), 900 µl NaCl solution (5 M) and 6 ml 5% (v/v) Nonidet P40 (NP40) in 22 ml nuclease-free water supplemented with 1% (v/v) protease- and 10% (v/v) phosphatase inhibitors. Cells were scraped off using a pre-chilled plastic cell scraper and were collected into pre-chilled 1.5 ml Eppendorf tubes. Harvested cells were rotated for 20 minutes at 4 °C and later centrifuged at maximum speed for 10 minutes at 4 °C. The supernatant was transferred into fresh tubes and was precleared by 30 minutes incubation with rotation with Gammabind Plus Sepharose$^{®}$ (50 µl per ml supernatant) at 4 °C, followed by 20 minutes centrifugation at

10 000 rpm at 4 °C. The supernatant was collected into a fresh tube and the protein concentration of the precleared supernatant was determined by NanoDrop™ ND-1000 UV/Vis Spectrophotometer (NanoDrop Technologies). Depending on the expression levels of the target protein, 5-10 mg of total cell protein lysate per immunoprecipitation (IP) reaction were incubated with respective antibody (at recommended dilution by manufacturer) and rotated at 4 °C for 2 hours. This was followed by addition of 35 µl of protein agarose beads (protein A or G based on their binding affinity for the antibody) and incubation with rotation at 4 °C overnight. On the following day, samples were centrifuged at 5000 rpm for 5 minutes at 4 °C. The supernatant (post-IP sample) was transferred into fresh pre-chilled tubes. The pellet was washed at least 3 times with 500 µl of above-mentioned lysis buffer (NP40 lysis buffer). To elute the proteins from the beads, the beads were mixed with IP sample buffer and were boiled at 95 °C for 6 minutes. The beads were precipitated by centrifugation at 7500 rpm for 8 minutes at 21 °C. The supernatant (IP sample) was transferred into a fresh tube at RT and proteins were separated on SDS-PAGE as described earlier (See 2.2.21). All steps were performed on ice and all rotations were done at medium speed.

2.2.25 Expression profiling

Gene expression profiling of CaMCaA cells was conducted in collaboration with Prof. Dr. med. Norbert Gretz and by Dr. Carolina De La Torre at the Medical Faculty Mannheim of the University Heidelberg (Mannheim, Germany). For that, PRSS23^{WT} and PRSS23^{P230A} expressing CaMCaA clones as well as CaMCaA parental control cells were cultured (300 000 cells per 10-cm dish). At 80-90% of confluency, RNA was isolated from total cell lysates using the NucleoSpin RNA kit (MACHEREY-NAGEL) following the instructions from the manufacturer. Capillary electrophoresis was performed to test the quality of RNA using an Agilent 2100 bioanalyzer (Agilent).

Gene expression profiling of CaMCaA control, two wild-type clones (WT16 and WT17), and two mutant clones (MUT3 and MUT4) was done using the GeneChip™ Mouse Gene 2.0 ST Arrays (Affymetrix). Biotinylated antisense complementary RNA (cRNA) was prepared using the GeneChip® WT Plus Reagent Kit (Affymetrix) and the GeneChip® Hybridization, Wash and Stain Kit (Affymetrix), following the Affymetrix standard labelling protocol. Next, On-chip hybridization was performed using a GeneChip Hybridization oven 640, then dyed in the GeneChip Fluidics Station 450 and thereafter scanned with a GeneChip Scanner 3000 (all the equipment were from Affymetrix).

The arrays were annotated using a Custom CDF Version 22 with ENTREZ based gene definitions [122]. Quantile normalization and RMA background correction were applied to the raw fluorescence intensity values and differentially expressed genes were identified by performing ANOVA using a commercial software package SAS JMP11 Genomics, version 7, from SAS (SAS Institute, Cary, NC, USA). The level of significance was set as false positive rate of a=0.05 with FDR correction.

Gene Set Enrichment Analysis (GSEA) was performed to evaluate the statistical significant bias of defined gene sets in their distribution within a ranked gene list using the R software-packages EnrichmentBrowser. [123] Pathways associated with cellular functions were obtained from public external databases (KEGG, http://www.genome.jp/kegg).

2.2.26 Mass spectrometry

In order to identify the interacting proteins of PRSS23^{WT} and PRSS23^{P230A}, immunoprecipitation of the C-terminally FLAG-tagged PRSS23 variants expressed in CaMCaA cells was performed using the monoclonal ANTI-FLAG® M2 antibody (Sigma-Aldrich). For detailed description refer to 2.2.24. Mass spectrometry (MS) analysis of the IP samples was done in collaboration with Prof. Dr. med. Darjus Tschaharganeh and by Dr. Dominic Helm and Martin Schneider at the Genomics and Proteomics Core Facility of The German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ).

Efficacy of immunoprecipitation was validated in IP samples before mass spectrometric analysis. First, IP samples were run on SDS-PAGE and the gel was washed three times with Millipore water (each time for 10 minutes). Then the gel was incubated with Coomassie Brilliant Blue dye for 2 hours at RT. Afterwards, the gel was washed with Millipore water overnight. All incubation steps were done under mild agitation. The efficacy of immunoprecipitation was confirmed by presence and intensity of other blue stained bands in addition to the band belonging to C-terminally FLAG-tagged PRSS23 in IP lanes. Second, IP samples were analyzed by Western blot for detection of C-terminally FLAG-tagged PRSS23 using the DYKDDDDK Tag antibody (Cell Signaling). For detailed description refer to 2.2.21.

For MS sample preparation, IP samples have been applied to SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) for a 0.5 cm run. Following this brief fractionation and Coomassie Blue staining, gel pieces corresponding to each sample were cut out individually and used for In-Gel Trypsin digestion of proteins according to Shevchenko et al. [124] on a DigestPro MSi robotic system (INTAVIS Bioanalytical Instruments AG).

The analysis was conducted on an Ultimate 3000 UPLC system (Thermo Fisher Scientific) coupled to an Orbitrap Exploris 480 mass spectrometer and was adapted to complexity of sample or dynamic range. Prior to the analysis, desalting of peptides on a Thermo Fisher Scientific trap column cartridge (Acclaim PepMap300 C18, 5μm, 300Å wide pore) was performed using 30 μl/min flow of 0.05% Trifluoroacetic acid (TFA) in water for three minutes. The multiple step gradient analysis was carried out on a nanoEase MZ Peptide analytical column (300Å, 1.7 μm, 75 μm × 200 mm, Waters) using two solvents; one consisting of 0.1% formic acid in water and the other consisting of 0.1% formic acid in acetonitrile. The concentration of the latter solvent (0.1% formic acid in acetonitrile) was linearly elevated from 2% to 38% for a differing time according to which analysis procedure was taken, followed by a quick escalation to 95%. After two minutes, the concentration was decreased back to 2% which was followed by a 10-minute equilibration. The peptides eluted were analyzed in the mass spectrometer using the standard method Data Dependent Acquisition (DDA). Full scan with 60k mass resolution (380-1400 m/z, 300% AGC target, 45 ms maxIT) and subsequent 1.5 seconds of MS/MS scans were performed. Peptide molecules with the mass range of 1.4 m/z were isolated and fragmented using 26% NCE (Normalized Collision Energy). Fragment spectra were acquired at 15k resolution (100% AGC target, 54 ms). Eluting molecules which were unassigned or singly charged, were excluded from fragmentation with a dynamic exclusion of 35 seconds.

Data analysis was done in MaxQuant (version 1.6.14.0) [125] by employing an organism-specific database derived from Uniprot.org. Identification FDR cutoffs were 0.01 on both peptide and protein levels. Peptide identifications were transferred across Raw files in respect to retention time and m/z via the Match between runs option.

The R package limma was employed for data analysis [126] and In-house compiled R-scripts were used to filter and visualize the MaxQuant output files. Protein quantification was done using the intensity-based absolute quantification (iBAQ). [127] Proteins with non-zero intensity values in all three replicates of at least one condition were included in the analysis. In total 1873 proteins could have been quantified from which 541 were quantified in all samples.

Data were normalized according to Variance Stabilization Normalization (VSN). [128] Values showing a presence absence pattern in between conditions were imputed via down shifted normal distribution. [129] Values showing no complete absence in one condition were imputed using the missForest R package. [130] The Benjamini–Hochberg adjustment for

multiple testing was applied to p-values. [131] Bioinformatics and statistical analysis of MS data was performed by Martin Schneider (Genomics and Proteomics Core Facility of DKFZ, Heidelberg).

2.2.27 Electron microscopy

Electron microscopy (EM) analysis of the CaMCaA cells expressing PRSS23^{WT} and PRSS23^{P230A} was performed in collaboration with Dr. Ingrid Haußer-Siller (Institute of Pathology, University of Heidelberg). Cells were cultured on sterile cover slips for 72 hours. On day 3, cells were 95% confluent and were fixed with pre-chilled 3% (v/v) Glutaraldehyde in Cacodylate buffer (0.1 M) for 2 hours. Cells were then washed 3 times (each 10 minutes) with the Cacodylate buffer (0.1 M, pH=7.4) followed by 30 minutes fixation with 1% (v/v) Osmium tetroxide in Cacodylate buffer (0.1 M) and 30 minutes of another wash step. Next, cells were treated with 5 minutes incubations in 30-100% (v/v) Ethanol solutions (gradually increasing the concentration) with exception of 10 minutes incubation in 80% and 90% solutions. After an extra 15 minutes incubation in 100% Ethanol, cover slips were dipped in Propylene oxide and placed on a capsule filled with Epon Araldite. The samples were polymerized over night at approximately 60°C. The following day, the cells were stuck to the Epon block and the coverslips were removed by heat and razor blades. The block was then trimmed with a Leica EM TRIM 2. Semithin section (900 nm) and ultrathin sections (70-90 nm) were cut with a Diatome glass knife and a Reichert Ultracut E ultramicrotome. The ultrathin sections were put on a 200 Mesh Copper Grid (Plano). Samples were contrasted with Uranyl acetate for 5 minutes and rinsed first with water and Lead citrate for 2 minutes and with water again. The samples were analyzed at 80 kV using a JEM 1400 electron microscope (JEOL) equipped with a F216 digital camera (TVIPS).

2.2.28 MYC activation assay

Nuclear extracts from 24 hours cultured CaMCaA PRSS23^{WT} and PRSS23^{P230A} expressing cells were prepared using the Nuclear Extract kit (Active Motif) following the instructions from the manufacturer. The protein concentration of samples was quantified using the Bradford assay. BSA standard solutions of 2, 1.5, 1.25, 1, 0.75, 0.5, 0.25, 0.125, 0.05 mg/ml were prepared in 1:5 water diluted complete lysis buffer (Nuclear extract kit, Active Motif). 5 µl of standards, blank (1:5 water diluted complete lysis buffer) and samples (1:5 diluted in water) were pipetted in duplicates into wells of a 96-well plate. 200 µl Bradford reagent was added to each well and incubated at RT protected from light for 8 minutes. Absorbance (OD)

was measured at 595 nm with the FLUOstar® Omega Microplate Reader (BMG LABTECH). A standard curve was established based on the concentration and the OD values of the standard samples using the MARS Data Analysis Software (BMG LABTECH). The protein concentration of samples was calculated according to the standard curve using the same software. The fractionation efficiency was validated by Western blotting analysis of nuclear and cytoplasmic fractions using nuclear and cytoplasmic markers (refer to Figure 18B). 5 µg of nuclear extracts from PRSS23 wild-type and PRSS23 mutant CaMCaA clones together with nuclear extracts from 24 hours cultured Jurkat cells (positive control for MYC activation provided in the kit) were used to measure the level of active MYC by using the TransAM™ c-Myc kit (Active Motif) following the instructions from the manufacturer. Three biological replicates were evaluated per clone. The absorbance was measured at 450 nm using the FLUOstar® Omega Microplate Reader (BMG LABTECH).

2.2.29 Flow cytometric analysis of cell cycle: propidium iodide staining

The distribution of the cell cycle phases in PRSS23^{WT} and PRSS23^{P230A} CaMCaA cells was investigated by flow cytometric analysis of DNA content. Cells (500 000 cells in 10-cm plate) were seeded in starvation medium (FBS-free CaMCaA complete medium) for 24 hours. On the following day, the starvation medium was replaced by complete medium, and the cells were further cultured for 48 hours. All the cells in culture plate were harvested by collecting the culture medium, the PBS used to wash the cells, and the trypsinized cells into a 15 ml falcon. Cells were then centrifuged at 1000 rpm for 5 minutes. The supernatant was discarded, and the cell pellet was washed with 1 ml PBS and later resuspended in 200 µl ice-cold 70% (v/v) Ethanol. Cells were fixed in Ethanol at -20°C for 1-2 hours followed by 5 minutes centrifuge at 1000 rpm. The supernatant was carefully removed, and the pellet was washed with 500 µl of PBS. Next, half of the cells were resuspended in 200 µl PBS as unstained control cells and the other half were resuspended in 200 µl of FxCycle™ PI/RNase Staining Solution (Invitrogen) followed by 20 minutes incubation at RT protected from light.

The unstained control cells and PI-stained cells were transferred (200 µl per well in triplicates) into a round-bottom 96-well plate and were analyzed using the Guava® easyCyte™ HT Flow Cytometer (Luminex). For cell cycle analysis, cells were gated based on size and then singularity to exclude cellular debris and cellular aggregates using the Guava® InCyte™ Software (Luminex). Then, cells were gated by the propidium iodide (PI) signal (red channel

at 695 nm) and the range of each cell cycle phase was defined on histograms and the percentage per peak was calculated by the software.

2.2.30 Computer modelling of the PRSS23 protein structure

The structural model of PRSS23 was developed by Prof. Giulia Rossetti (Forschungszentrum Jülich). The ab-initio modeling method was employed in the iTASSER server [132] and the Phyre2 [133]. The structures derived from both servers were highly similar to each other and comparable to the structure of other serine proteases.

2.2.31 Statistics

Statistical analysis of data was performed using Microsoft Excel, GraphPad Prism, and the R software as indicated. Statistical significance was calculated using a two-tailed student's t-test. Multiple comparisons were carried out either by one-way- or two-way ANOVA, followed by Tukey's multiple comparison correction or Fisher's Least Significant Difference (LSD) test. p-values less than 0.05, 0.01, and 0.001 are represented as *, **, and *** on the figures, respectively. Error bars represent either the standard deviation (SD) or standard error of the mean (SEM), as indicated in the respective figure legends.

3 RESULTS

3.1 Identification of clonally expanded driver mutations of human hepatocellular carcinoma

In order to identify clonally expanded driver mutations of human hepatocarcinogenesis, a single HCC case was selected and used for whole exome sequencing (WES). This particular HCC was characterized by nodule-in-nodule growth of a well-differentiated HCC in a premalignant HGDN (Figure 4). Following microdissection, DNA was isolated from gallbladder (GB), cirrhotic liver (LC), dysplastic nodule (DN), and HCC tissues and genetic variants were detected using next generation sequencing. In the scenario of step-wise hepatocarcinogenesis, alterations that are joint by the DN and the HCC represent clonally expanded alterations, while variants present in the DN part only may be related to malignant transformation. To select clonally expanded variants in the WES data, the ANNOVAR pipeline was employed by Dr. Robert Geffers (Genome Analytics, Helmholtz Centre for Infection Research Braunschweig). [116] Using this approach, 17 clonally expanded variants were identified from which 11 variants were present in both the DN and the HCC compartments. The remaining six were exclusively present in the HCC nodule (Table 12). All variants were validated by Sanger sequencing.

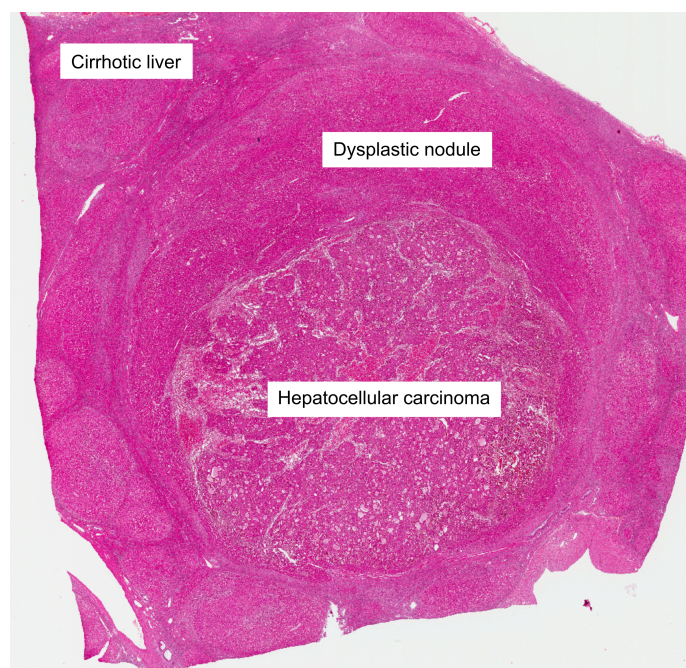**Figure 4. Histological section of the case analyzed by NGS.** The well-differentiated early HCC is surrounded by an HGDN leading to a nodule-in-nodule appearance. Whole exome sequencing was performed using microdissected cirrhotic liver tissue, DN, and HCC as well as normal gallbladder tissue from this patient to identify clonally expanded somatic variants driving malignant transformation from HGDN to early HCC.

Table 12. Summary of the clonally expanded variants.

Type	Gene	Amino acid change	Location	Somatic p-value	Polyphen-2	VF (DN)	VF (HCC)
SNP	NMNAT2	NM_015039:c.559C>G:p.R187G	DN and HCC	1.46E-13	0.991	26%	48%
SNP	MLL4	NM_014727:c.3730T>A:p.W1244R	DN and HCC	1.85E-12	0.999	32%	30%
SNP	HOGA1	NM_001134670:c.121C>T:p.P41S	HCC-specific	9.72E-10	1	ND	28%
SNP	DLK1	NM_003836:c.1012C>T:p.R338C	DN and HCC	1.19E-09	0.999	28%	24%
SNP	DIP2C	NM_014974:c.1612A>C:p.N538H	HCC-specific	1.81E-07	0.998	1%	22%
SNP	CLMP	NM_024769:c.860C>T:p.P287L	HCC-specific	3.84E-07	0.999	ND	20%
SNP	KCNQ2	NM_004518:c.478C>T:p.R160W	DN and HCC	5.90E-07	1	32%	36%
SNP	MAP3K4	NM_005922:c.547G>T:p.D183Y	DN and HCC	1.16E-06	0.999	37%	9%
SNP	RTN4IP1	NM_032730:c.653G>A:p.G218D	HCC-specific	3.14E-06	1	ND	38%
SNP	**PRSS23**	**NM_007173:c.688C>G:p.P230A**	**DN and HCC**	**4.04E-06**	**0.999**	**47%**	**32%**
SNP	TMPRSS15	NM_002772:c.1355T>A:p.F452Y	DN and HCC	2.08E-05	0.995	55%	14%
SNP	VWA5A	NM_014622:c.97G>A:p.V33M	DN and HCC	2.46E-05	0.998	19%	26%
SNP	PIK3CG	NM_002649:c.2553G>T:p.M851I	HCC-specific	0.0011	0.997	ND	29%
SNP	LAMA1	NM_005559:c.4531T>C:p.C1511R	HCC-specific	0.0053	1	ND	26%
SNP	FZD4	NM_012193:c.931T>G:p.F311V	DN and HCC	0.0118	0.993	27%	33%
SNP	FAAP100	NM_025161:c.2207T>G:p.L736R	DN and HCC	0.013	0.87	27%	30%
InDel	FBN2	NM_001999:exon54:c.6802_6802delinsC-	DN and HCC	6.21E-07	-	ND	29%

Abbreviations: DN, dysplastic nodule; HCC, hepatocellular carcinoma; InDel, insertion/deletion; SNP, single nucleotide polymorphism; delins, Deletion/Insertion; VF, variant frequency; ND, not detected.

3.2 Identification of tumor suppressive genes involved in driving malignant transformation of HCC

To validate the tumor promoting potential of the identified clonally expanded variants, *in vivo* RNAi screening was performed in collaboration with Prof. Dr. Lars Zender and his research group (University of Tübingen) using a mosaic mouse model of liver cancer (Figure 5). For this aim, a miR30-based shRNA library consisting of three independent shRNAs targeting each of the 17 candidate genes (in total 51 shRNAs) was designed. HTVI was used to deliver the shRNA library together with a *Myc*-Myr*AKT1* transposon vector into the liver of *Trp53* heterozygous mice. The sleeping beauty transposon system was employed for stable expression of the oncogenes MYC and constitutively active myristoylated AKT1 as well as the shRNA library in transformed hepatocytes. GFP expression was used to non-invasively monitor tumor formation. In this scenario, shRNAs depleting the expression of tumor suppressor genes provide a growth advantage to cancer cells and will thus be enriched within the HCC nodules. Overall, 13 HCC nodules were dissected, DNA was isolated, and the expressed shRNAs were detected and quantified using NGS (Figure 5A and 5B). Four candidates (*Vwa5a*, *Clmp*, *Dlk1*, and *Prss23*) were selected for further analyses. *Vwa5a* was

selected as it was targeted by the far most enriched shRNA. The remaining three genes were chosen since three different shRNAs targeting each of these candidates were found enriched in different murine HCC nodules, corroborating their role in driving malignant transformation of murine hepatocytes (Figure 5C and 5D).
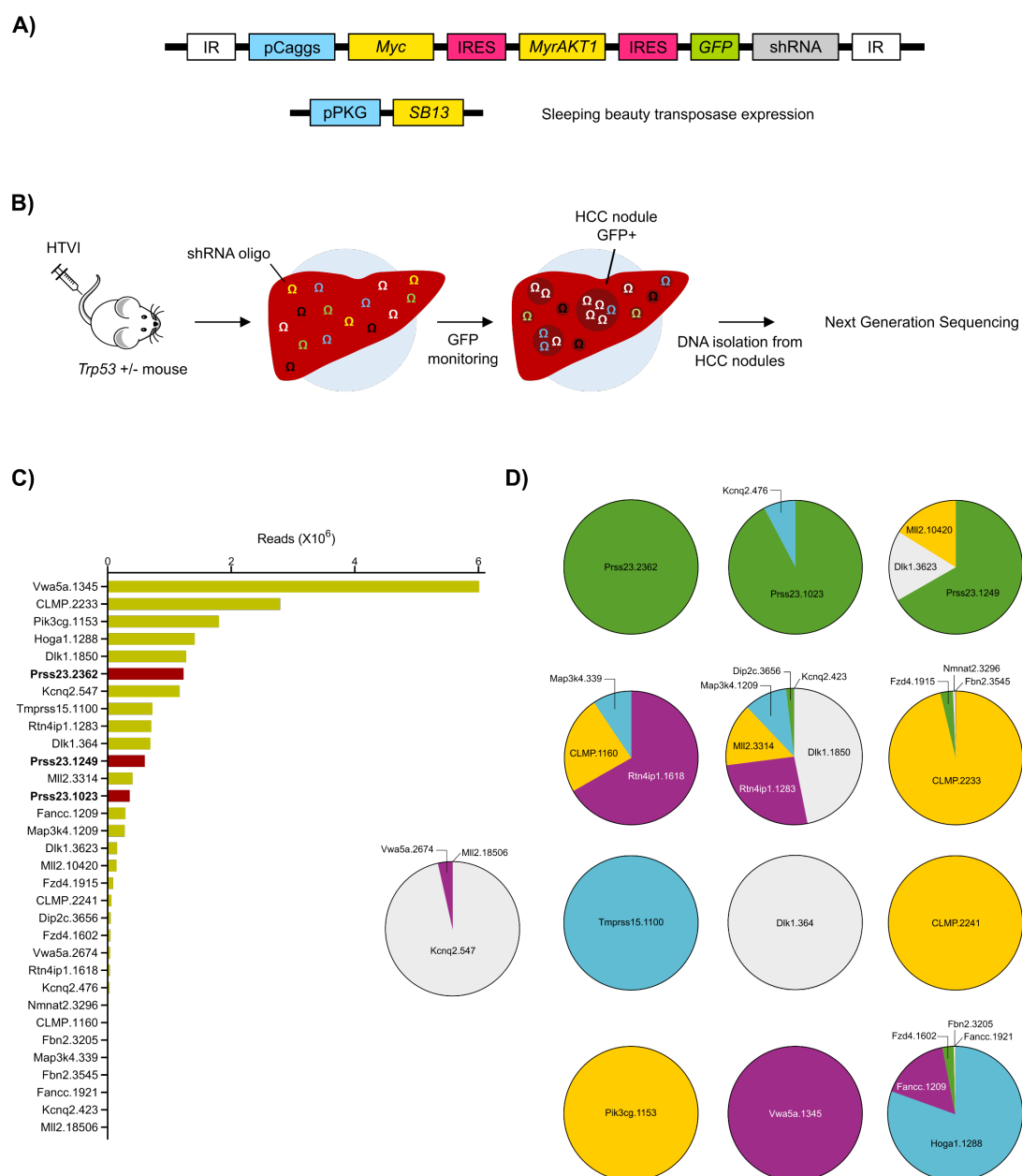
Figure 5. Identification of tumor suppressive genes using *in vivo* RNAi screening in mosaic mouse model of HCC. A) Transposon vectors harboring *Myc*-Myr*AKT1*-*GFP*-*shRNA* library and sleeping beauty (SB) transposase B) were co-delivered to liver of *Trp53* heterozygous mice using HTVI. Tumor growth was monitored non-invasively based on GFP expression of successfully transformed, tumor-forming hepatocytes. NGS was performed using nucleic acids isolated from individually dissected HCC nodules. C) Expression level of the shRNA detected in murine HCC nodules. D) Distribution of the expressed shRNAs in the 13 independent HCC nodules analyzed.

3.3 Validation of tumor promoting potential of the identified mutations in development of HCC

For single gene verification of the variants scored by the shRNA library screening, either the wild-type or the mutant variant of each of the candidate genes was stably expressed in p53 heterozygous mice, again using the sleeping beauty transposon system. In brief, a CaMIG plasmid expressing *Myc* and *GFP* was co-delivered with a pCaggs vector carrying the gene of interest using HTVI (Figure 6A). Expression of *Myc* alone does not lead to HCC development in *Trp53* heterozygous mouse. However, 70% of mice (N=7/10) expressing the mutant *Prss23*P230A and 12.5% of mice (N=1/8) expressing the wild-type *Prss23* gene developed HCC (Figure 6B and 6C). This data confirms that the mutation c.686C>G of the *Prss23* gene (BC018517) leading to replacement of proline by alanine at position 230 of the PRSS23 protein is indeed a driver of malignant transformation in MYC-expressing *Trp53* heterozygous murine hepatocytes.
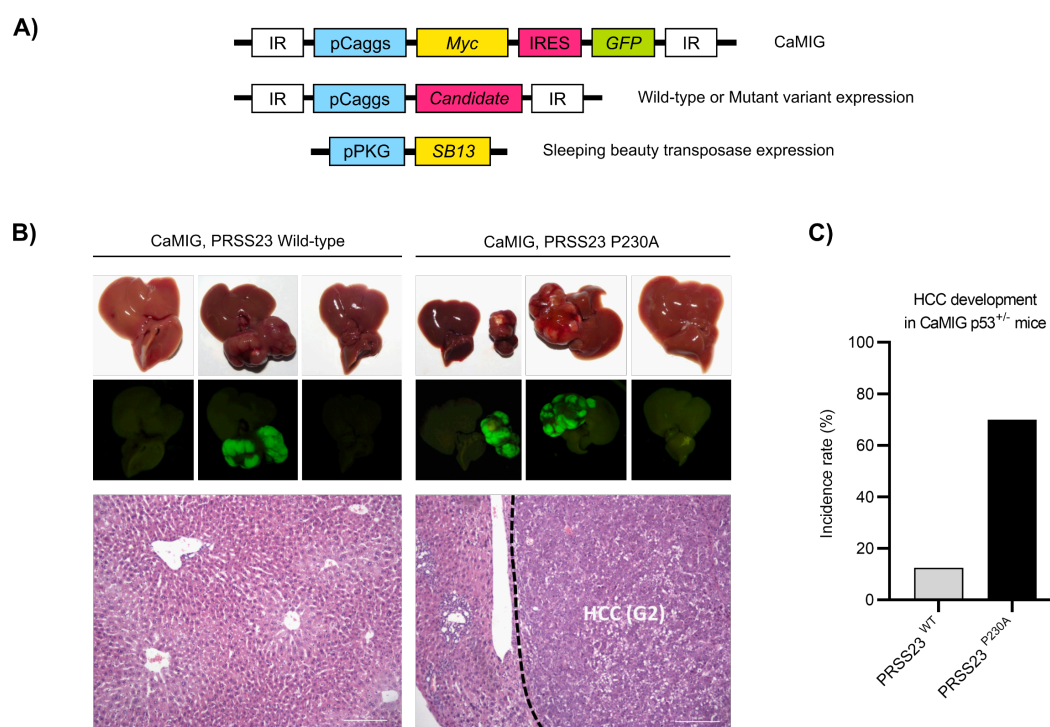
Figure 6. *In vivo* validation of the individual genetic variants identified using MYC-induced mosaic mouse model of HCC. A) A transposon vector expressing either the wild-type or the mutant variant of the candidate gene was co-injected together with CaMIG (*Myc-GFP*) and the sleeping beauty transposase carrying vectors into *Trp53* heterozygous mice. B and C) After 6 months, 7 out of 10 mice expressing *Prss23*P230A variant and 1 out of 8 mice expressing the wild-type *Prss23* gene developed HCC.

3.4 Structural modelling of PRSS23

To understand the effect of the P230A mutation on the PRSS23 protein structure, a computational 3D model was developed by Prof. Giulia Rossetti (Forschungszentrum Jülich). The comparison of this PRSS23 model with other serine proteases implies that the catalytic triad resides in residues H175-S341-D246. Interestingly, proline 230 is located in close vicinity to this catalytic triad (Figure 7). The model suggests that the proline to alanine substitution at amino acid 230 may be close enough to establish a salt bridge with D246 (aspartic acid) and may thus affect the catalytic activity of PRSS23.
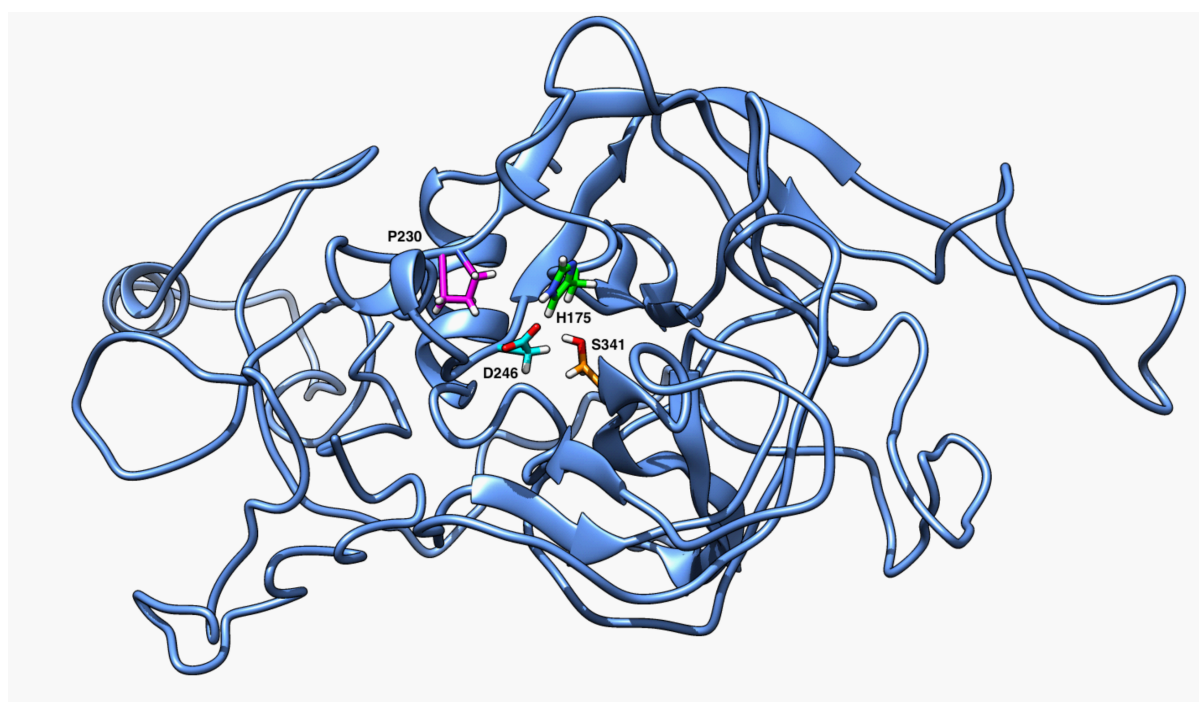
Figure 7. Model of PRSS23 tertiary structure in ribbon representation. Three residues of the catalytic site H175-S341-D246 are respectively marked in green, orange, and cyan licorice representation. The residue P230 harboring the mutation identified in this study is shown in magenta licorice representation.

3.5 Establishment of isogenic cell lines overexpressing the wild-type or the mutant variant of PRSS23

In order to study the function of wild-type and mutant PRSS23 in more details and to investigate their role in promoting HCC development, isogenic cell lines overexpressing wild-type or mutant PRSS23 were established. For this purpose, the primary murine HCC cell line CaMCaA, which was established from a mosaic mouse HCC induced by MYC and AKT1, was transfected with a transposon vector overexpressing either the wild-type or the P230A mutant of *Prss23*. The expression of wild-type and mutant PRSS23 was validated at mRNA and

protein levels in the derived single cell clones (Figure 8A and 8B). Due to the lack of antibodies reliably detecting PRSS23, I established isogenic CaMCaA clones expressing C-terminally FLAG-tagged PRSS23 variants which allowed for indirect detection of the PRSS23 protein using anti-FLAG tag antibodies (Figure 8B and 8C). Establishment of CaMCaA clones expressing N-terminally FLAG-tagged PRSS23 was not successful, likely due to post-translational cleavage of the N-terminal activation domain of PRSS23 during zymogen activation. [86]
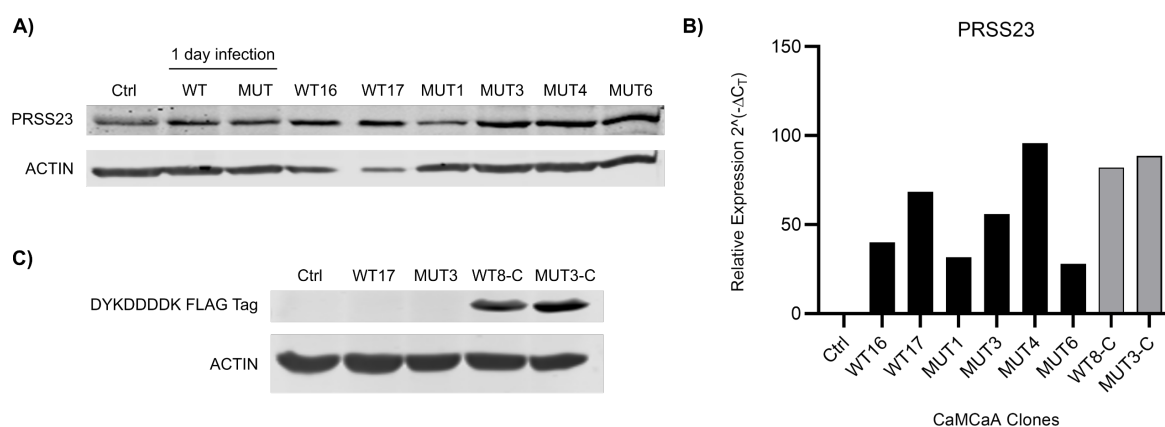
Figure 8. Establishment of primary murine HCC cell lines expressing PRSS23 WT/P230A variants. The primary murine HCC cell line (CaMCaA) derived from a MYC-MyrAKT1 induced murine HCC was transfected with a transposon vector expressing either the *Prss23* wildtype or the P230A variant under control of an EF1-alpha promoter. Expression of PRSS23 wild-type and P230A mutant was validated at A) protein level (actin is detected as loading control) and B) mRNA level in WT and mutant clones as well as PRSS23 WT/P230A C-terminal FLAG-tagged CaMCaA clones. The data shown represent the relative expression compared to the parental CaMCaA control. N=1. Abbreviations: Ctrl, parental CaMCaA cells; WT, CaMCaA clones expressing PRSS23 wild-type; MUT, CaMCaA clones expressing PRSS23[P230A]; WT-C and MUT-C, clones expressing the respective C-terminal FLAG-tagged PRSS23 variants; the numbers included in the labelling of the clones indicate different single clones. Establishment of WT and MUT untagged PRSS23 clones was conducted by Dr. Federico Pinna (Longerich lab).

3.6 Gene expression profiling of CaMCaA PRSS23 WT/P230A Mutant

To understand and explain the phenotypic and functional differences observed between PRSS23[WT] and PRSS23[P230A] expressing HCC cells, comparative gene expression profiling was performed in collaboration with Prof. Dr. med. Norbert Gretz and by Dr. Carolina De La Torre at the Medical Faculty Mannheim of the University Heidelberg (Mannheim, Germany). Gene Set Enrichment Analysis (GSEA) revealed a significant upregulation of cell cycle and growth associated pathways including replication and nucleotide metabolism, transcription and translation in mutant compared to wild type clones, whereas metabolic pathways in particular

carbohydrate, glycan, amino acid, and lipid metabolism were significantly downregulated in the PRSS23^{P230A} mutant clones. Among signal transduction pathways, TGF-β and Hippo signaling were upregulated in mutant clones, while mTOR, HIF-1, FOXO, phosphatidylinositol, and sphingolipid signaling pathways were downregulated compared to wild-type clones. Having a closer look at the enriched pathways and processes, it appeared that transport and catabolism namely lysosome, peroxisome, endocytosis, and autophagy were remarkably suppressed in PRSS23^{P230A} mutant CaMCaA clones compared to wildtype CaMCaA clones. Although both PRSS23^{P230A} and PRSS23^{WT} proteins were expressed in MYC-AKT1-driven HCC cells, the pathways analysis indicated that MYC signaling and MYC-dependent crosstalk were enhanced in the PRSS23^{P230A} expressing clones (Figure 9).

3.7 Expression of PRSS23^{P230A} promotes survival and proliferation of CaMCaA cells

The phenotype of PRSS23^{P230A} CaMCaA clones suggested an increased proliferation rate compared to the wild-type controls *in vitro*. This was evident on the one hand from the degree of confluency following 72 hours of culture and on the other hand from the more acidic pH of the culture medium as suggested by a change in medium color (Figure 10A). In order to scrutinize this phenotype, I measured the cell viability of CaMCaA clones (Figure 10B). As expected, PRSS23^{P230A} clones showed a significantly higher viability, proliferation rate, and growth capability compared to clones expressing the wild-type PRSS23 protein. Moreover, cell cycle analysis revealed that PRSS23^{P230A} expressing cells were enriched in the S and G2/M phases of the cell cycle in comparison to wild-type CaMCaA clones. On the other hand, about half of CaMCaA PRSS23^{WT} cells were accumulated in G0/G1 phase compared to ~30% in their mutant counterparts (Figure 10C). These data suggested that a greater fraction of CaMCaA PRSS23^{P230A} expressing cells was actively proliferating at 72 hours of culture. In line with the cell cycle analysis, expression profiling revealed a consistent differential regulation of cell cycle associated genes in CaMCaA PRSS23^{P230A} expressing cells compared to wild-type clones. Congruently, gene markers of G0/G1 phase including Cyclin D1/2 (*Ccnd1/2*), Histone Deacetylase 1 (*Hdac1*), and Checkpoint kinase 2 (*Chek2*) showed higher expression levels in CaMCaA PRSS23^{WT} clones in comparison to markers of S and G2/M phases of cell cycle, Cyclin E, A, and B, and Cyclin D kinase 1/2 (*CDK1/2*), to name a few, which show elevated levels in CaMCaA PRSS23^{P230A} clones (Figure 10D).
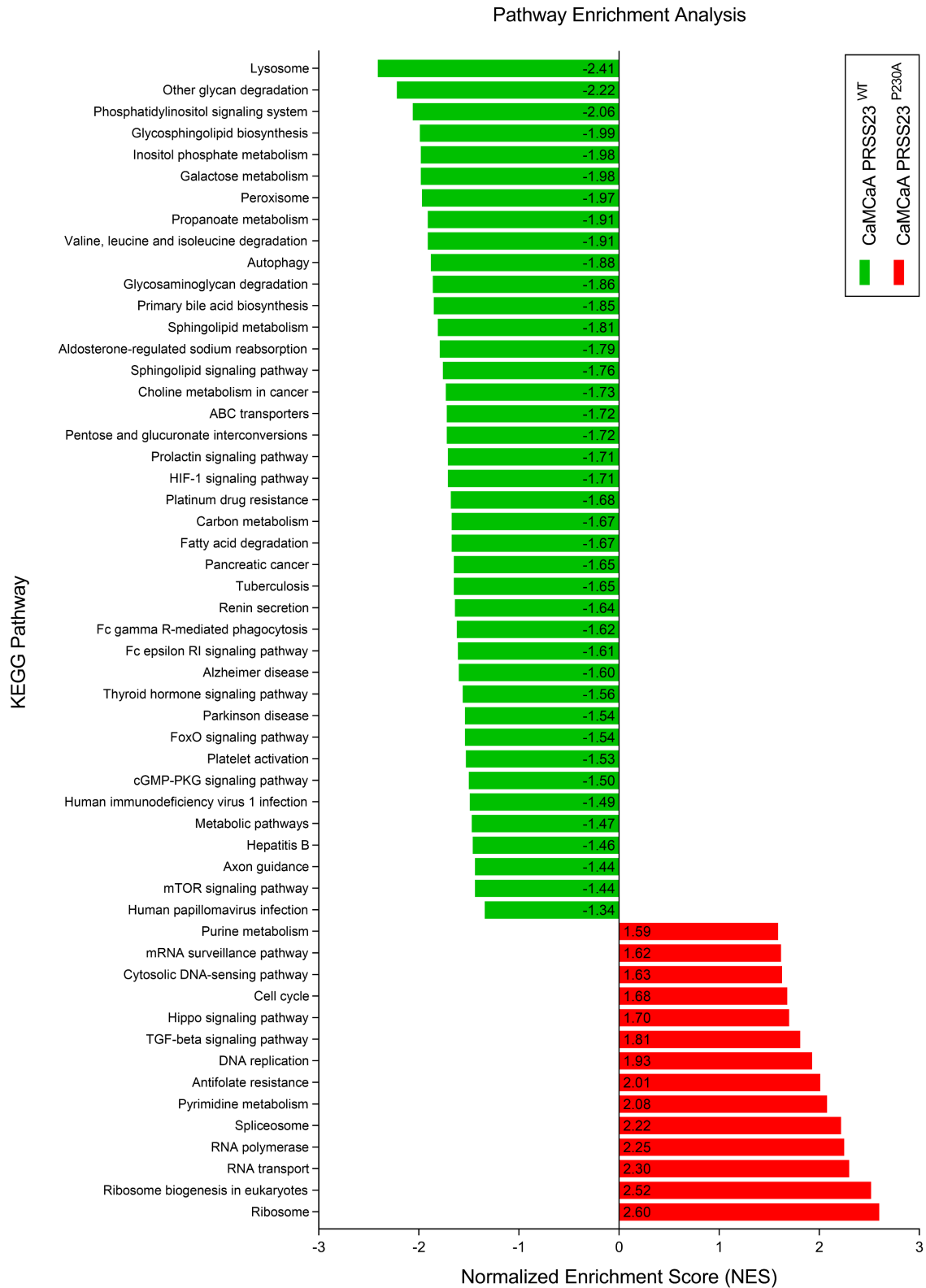
Figure 9. Differentially regulated signaling pathways and cellular processes between PRSS23[WT] and PRSS23[P230A] expressing CaMCaA cells. Positive NES values (red bars) represent enrichment in mutant clones and negative NES values (green bars) represent enrichment in wildtype clones. The threshold was set as logFC > 2 and adjusted p-value < 0.05. Number of replicates per clone = 3.
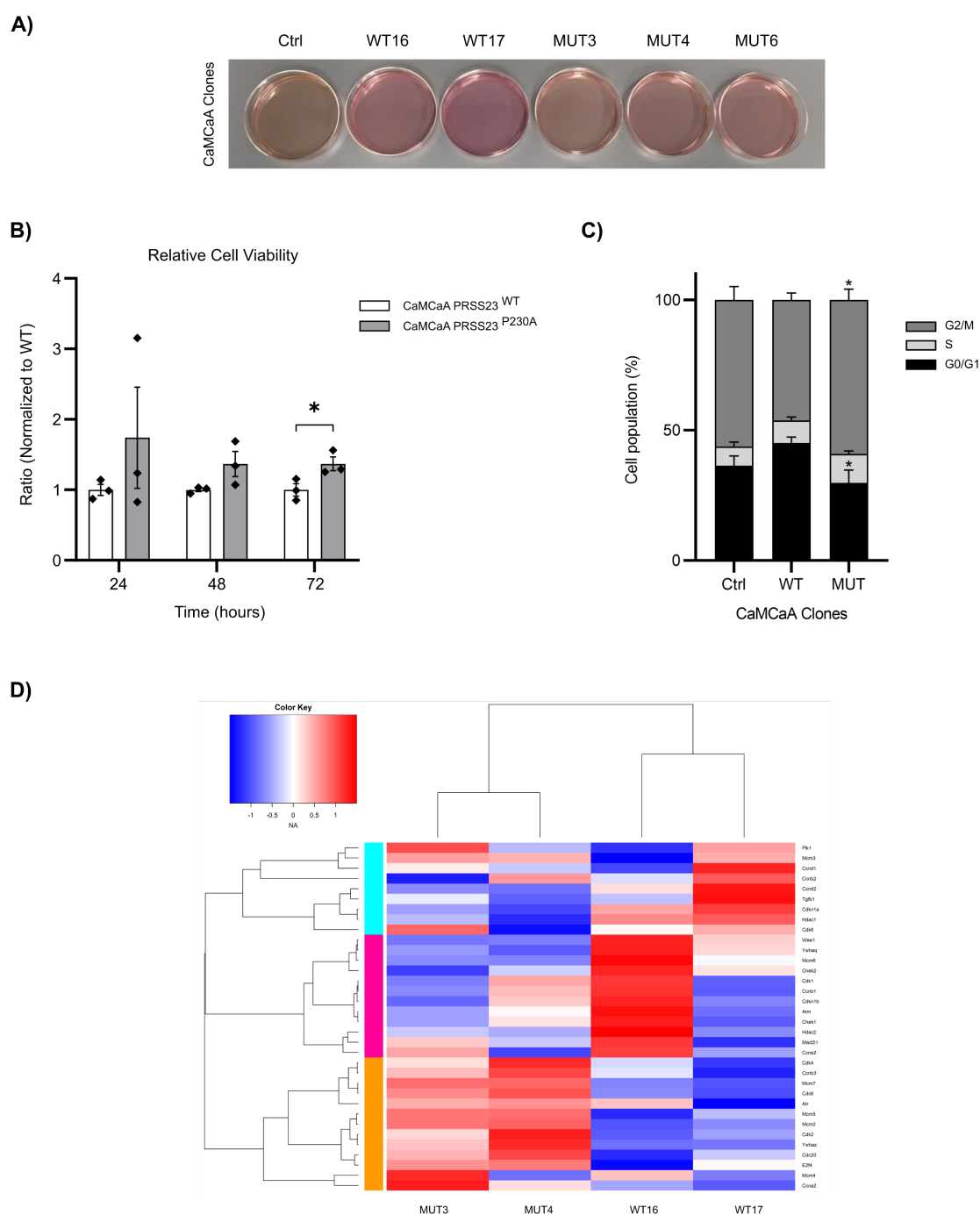
Figure 10. Expression of PRSS23^{P230A} enhances cell viability and proliferation capacity of CaMCaA cells. A) The medium of 72 hours cultured CaMCaA clones. The more acidic (yellowish) color is notable in parental control and mutant clones compared to the wild-type clones. B) The cell viability was significantly increased in PRSS23^{P230A} compared to PRSS23^{WT} clones at 72 hours of culture. Data are mean ± SEM of three independent experiments. Data are normalized to PRSS23^{WT} clone of respective timepoint and experiment. * P <0.05, unpaired two-tailed student's t-test. C) CaMCaA cells (500 000 cells per clone per 10-cm dish) were seeded and serum starved for 24 hours to achieve cell cycle synchronization. The cell cycle analysis was performed following 72 hours of culture. Data represent the mean ± SEM of three independent experiments. *P <0.05, two-way ANOVA followed by Tukey's multiple comparisons correction. D) Heatmap plot shows differential expression of cell cycle phase associated gene markers between PRSS23^{WT} and PRSS23^{P230A} expressing CaMCaA clones. Abbreviations: Ctrl, CaMCaA control; WT: CaMCaA PRSS23^{WT} expressing cells; MUT: CaMCaA PRSS23^{P230A} expressing cells.

3.8 Overexpression of PRSS23^{P230A} modules metabolic pathways in CaMCaA cells

To validate the deregulation of the selected metabolic pathways at protein levels in CaMCaA PRSS23^{WT} and PRSS23^{P230A} expressing cells, Western blot analysis of total cell lysate proteins extracted from CaMCaA cells at 48, 72, and 96 hours of culture was performed (Figure 11). This analysis confirmed the differential regulation of signaling pathways associated with cellular metabolism (mainly AMPK, COX-2, and PI3K-AKT-mTOR) between PRSS23^{WT} and PRSS23^{P230A} expressing CaMCaA cells. The higher levels of phosphorylated AMPK (pAMPK) and lower levels of COX-2 were consistently observed in wild-type clones compared to mutant clones over different timepoints. These results are in line with lower proliferative rates in wild-type clones and higher fractions of cells in G0/G1 phase of the cell cycle (Figure 10B and 10C, respectively).
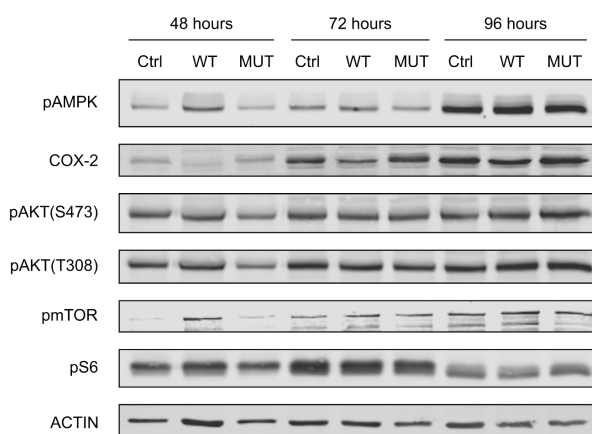
Figure 11. Expression of PRSS23^{P230A} modulates metabolic pathways in CaMCaA cells. Western blot was performed on protein isolated from total cell lysate of CaMCaA cells cultured (500 000 cells per 10-cm dish) for 48, 72, and 96 hours. Protein markers of metabolic associated signaling pathways including AMPK, PI3K-AKT-mTOR, and COX-2 signaling were detected. N = one representative experiment. Abbreviations: Ctrl, CaMCaA control; WT: CaMCaA PRSS23^{WT} expressing cells; MUT: CaMCaA PRSS23^{P230A} expressing cells.

3.9 Expression of PRSS23^{P230A} increases invasion capacity in CaMCaA cells

To investigate whether PRSS23^{P230A} affects the invasive properties of CaMCaA cells, I performed the transwell invasion assay using the Corning$^{®}$ BioCoat$^{™}$ Growth Factor Reduced Corning$^{®}$ Matrigel$^{®}$ Invasion Chamber (Corning). These analyses illustrated that PRSS23^{P230A} promotes invasion of CaMCaA cells compared to PRSS23^{WT} control cells (Figure 12A and 12B).
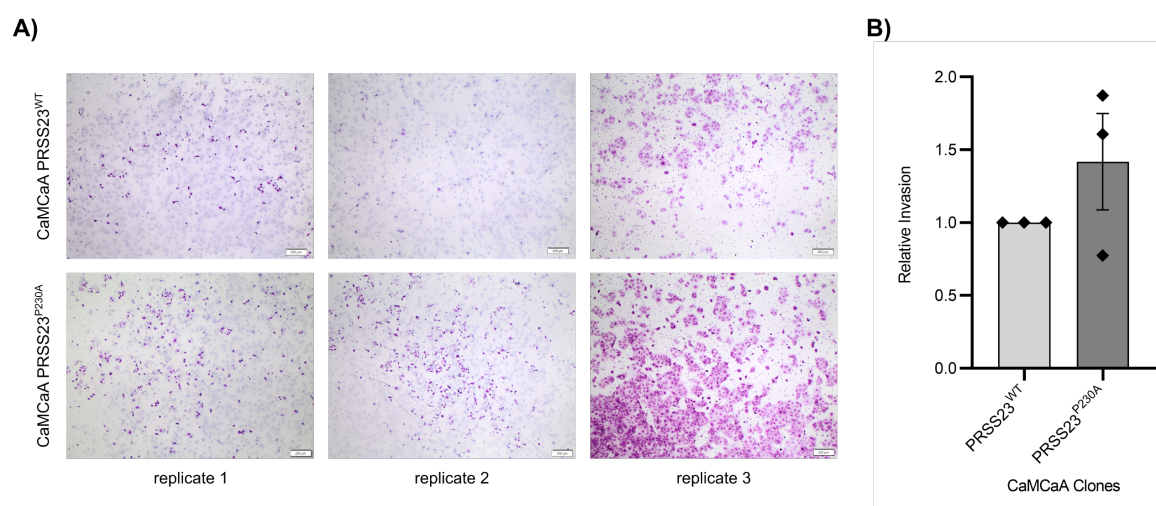
Figure 12. CaMCaA PRSS23^{P230A} expressing cells demonstrated higher cell invasion capacity compared to PRSS23^{WT} controls. CaMCaA cells expressing either the wildtype or P230A variant of PRSS23 were cultured in invasion chambers in serum-deprived medium for 24 hours. On the following day, non-invaded cells were removed from the upper side of the membrane and invaded cells were fixed to bottom side of the membrane for following microscopic analysis. A) Microscopic images of invaded PRSS23^{WT} and PRSS23^{P230A} expressing CaMCaA cells (one representative image per biological replicate). 4× magnification; Scale bar, 200 μm. B) Densitometric analysis of microscopic images was performed to measure the area occupied by invading cells. The total invaded area per field was used to compare the invasion capacity of CaMCaA PRSS23^{WT} and PRSS23^{P230A} expressing cells. Data are normalized to respective PRSS23^{WT} expressing control cells. Five fields per membrane and three biological replicates were included in the analysis. Unpaired two-sided student's t-test was used for statistical analysis.

3.10 P230A does not affect cellular localization of PRSS23 in CaMCaA cells

To unravel the molecular mechanisms by which PRSS23^{P230A} drives hepatocarcinogenesis, cellular localization of PRSS23^{WT} and PRSS23^{P230A} was investigated in CaMCaA cells. As PRSS23 belongs to the S1A family of serine proteases [87, 88], it seemed likely that PRSS23 constitutes a secreted protease engaged in ECM remodeling. This would explain the increased invasive capacity of PRSS23^{P230A} compared to PRSS23^{WT} cells (refer to 3.9). To substantiate this, I measured the concentration of the PRSS23 protein in the culture medium of different CaMCaA clones by ELISA. Surprisingly, the extracellular concentration of PRSS23 was very low and no difference could be observed between wild-type and mutant clones (Figure 13A).

Next, I performed subcellular protein fractionation and detected protein levels of PRSS23^{WT} and PRSS23^{P230A} in mitochondria, cytoplasm, and the nucleus of all CaMCaA clones analyzed (Figure 13B and 13C). Surprisingly, PRSS23 was highly expressed in the mitochondrial fraction of both wild-type and mutant clones. The PRSS23 protein levels were

similar between wild-type and mutant clones suggesting that the P230A mutation does not affect the cellular localization and distribution of the protein in CaMCaA cells and the functional differences between wild-type and P230A clones may either be explained by the enzymatic activity of the protein or its binding affinity to other protein.
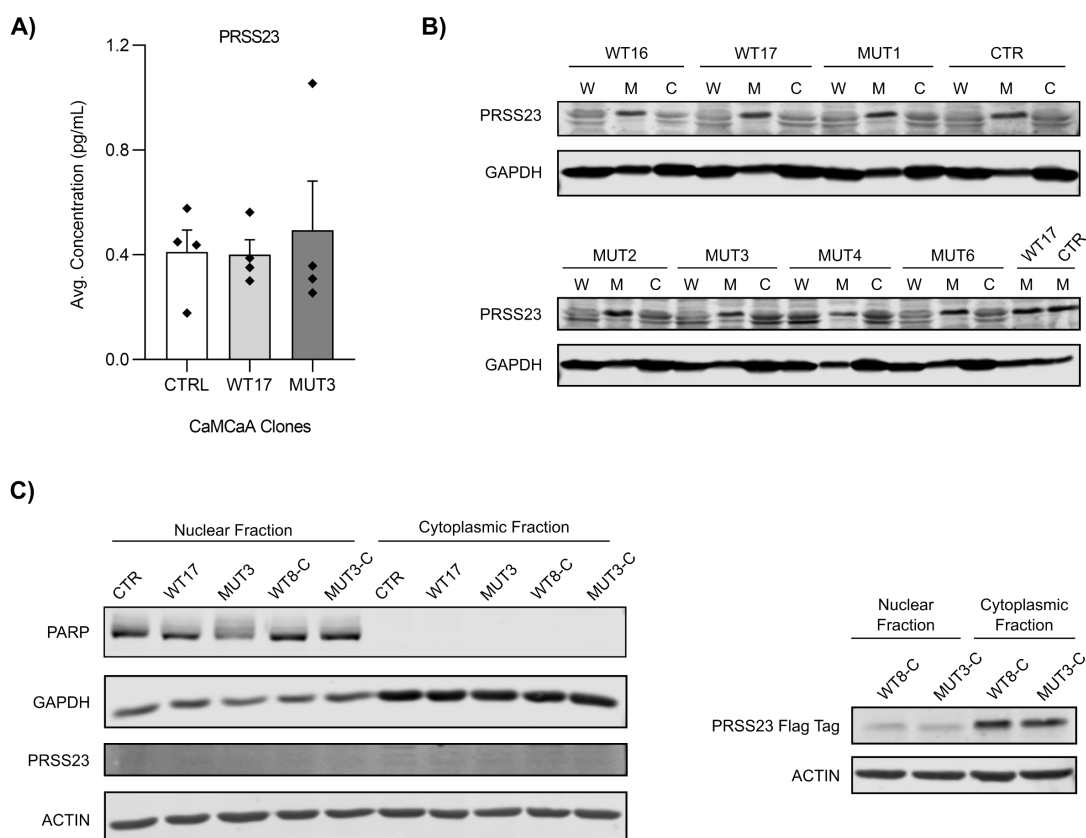
Figure 13. Cellular compartmentalization of PRSS23 in CaMCaA cells. A) ELISA of CaMCaA medium following 72 hours culture. Although there is a slight increase in mutant compared to wild-type clones, the difference was not statistically significant. Data represents means ± SEM of four independent experiments, each done in triplicates. One-way ANOVA with Tukey's multiple comparisons test was used for statistical analysis. B) Mitochondria and C) nuclear, and cytoplasmic protein fractions were extracted from CaMCaA clones as described in 2.2.21. Western blot analysis revealed an enrichment of PRSS23 both in wild-type as in P230A mutant clones compared to the whole cell lysate and the cytoplasmic compartment. Western blot analysis of PRSS23 protein levels in wild-type and mutant clones with or without a C-terminal FLAG tag reveals only a very weak band using a commercial PRSS23 antibody. However, the presence of PRSS23^{WT} and PRSS23^{P230A} protein in the nucleus and at higher levels also in the cytoplasm was validated using the FLAG tag antibody. Abbreviations: W, whole cell lysate; M, mitochondrial fraction; C, cytoplasmic fraction; WT, PRSS23 wild-type; MUT, PRSS23^{P230A}; WT-C, C-terminally FLAG-tagged PRSS23 wild-type; MUT-C, C-terminally FLAG-tagged PRSS23^{P230A}.

3.11 PRSS23^{P230A} promotes ATP production in CaMCaA cells

The expression profiling analysis described above showed that pathways involved in cellular metabolism and energy homeostasis including AMPK, PI3K-AKT-mTOR, and HIF-1 signaling were differentially activated between PRSS23^{P230A} and PRSS23^{WT} expressing CaMCaA cells (refer to 3.6; Figure 9). Additionally, PRSS23 was highly enriched in the mitochondria fraction of these cells (refer to 3.10; Figure 13B). Thus, it seemed reasonable to investigate the cellular energetics of different CaMCaA clones.

The ATP production rate in both glycolysis and oxidative phosphorylation was quantified under normal oxygen concentration and hypoxia using a Seahorse Analyzer (Seahorse Real-Time ATP rate assay, Agilent). The ATP production rate was significantly higher in PRSS23^{P230A} compared to PRSS23^{WT} clones under normoxia, which was seen both for glycolysis and mitochondrial respiration (Figure 14A-C). In particular, the mitochondrial ATP production rate was accelerated in PRSS23^{P230A} compared to PRSS23^{WT} clones (Figure 14B). Interestingly, incubation of the HCC cells with DMOG (chemical inducer of hypoxia) for two hours prior to analysis significantly reduced the mitochondrial ATP production rate in PRSS23^{P230A} and elevated the glycolytic rate in PRSS23^{WT} clones, which substantially abated the differences in cellular energetics between wild-type and mutant clones. Of note, PRSS23^{WT} cells were able to switch from mitochondrial respiration to anaerobic glycolysis following the induction of hypoxia by DMOG. However, PRSS23^{P230A} expressing CaMCaA cells failed to undergo this adaptation step (Figures 14A-C).

Moreover, PRSS23^{P230A} expressing CaMCaA cells require a higher oxygen consumption rate (OCR) compared to PRSS23^{WT} cells indicating an increased cellular respiration rate; a phenotype being reversed under hypoxic conditions (Figure 14D and 14E). Noteworthy, the differences in OCR between wild-type and mutant clones relies on the activity of ATP synthase and is most likely not linked to electron transfer chain (ETC) complexes I and III of the mitochondria as the basal OCR differences between wild-type and mutant clones is already completely abolished following inhibition of ATP synthase by oligomycin. In addition, it does not show a further reduction following rotenone plus antimycin A (Rot/AA) treatment, which inhibit the mitochondrial ETC (Figures 14D and 14E).

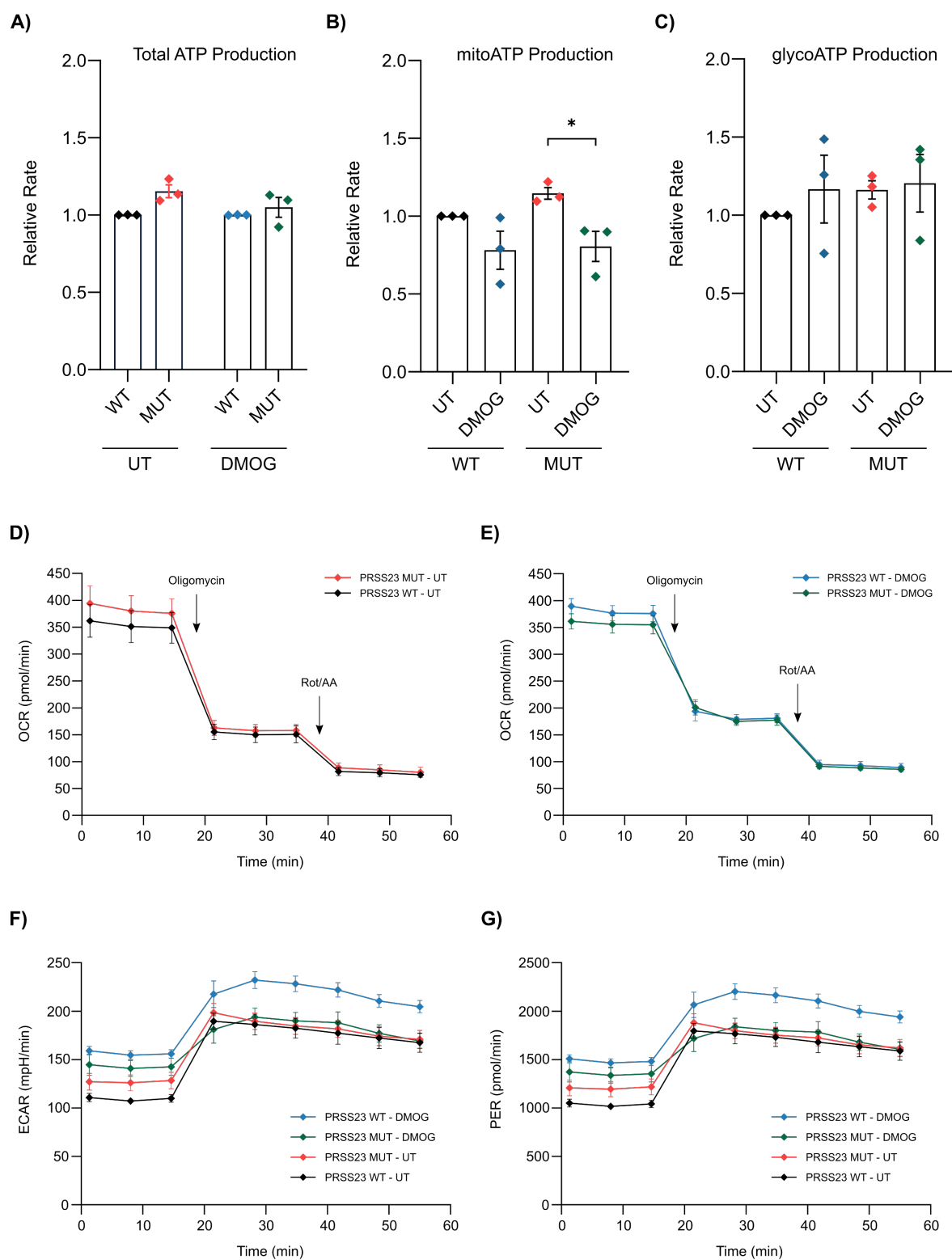**Figure 14. PRSS23^{P230A} improves ATP production rate in CaMCaA cells.** The ATP production rate of glycolysis and oxidative phosphorylation was measured in PRSS23^{WT} and PRSS23^{P230A} expressing CaMCaA clones under normoxic and hypoxic conditions using the Seahorse Real-Time ATP rate assay (Agilent). A) Total ATP production rate of CaMCaA PRSS23^{P230A} was higher compared to CaMCaA PRSS23^{WT} overexpressing

cells under normoxia and the observed difference was abated following the treatment of cells with DMOG (chemical inducer of hypoxia). B) mitoATP production rate of PRSS23^{P230A} expressing CaMCaA cells was higher compared to the CaMCaA cells expressing the wildtype protein. Following DMOG treatment, both clones showed reduced mitochondrial respiratory rate, but this reduction was significantly more present in the mutant clone. C) In a similar manner to mitoATP, the rate of glycoATP production was higher in the mutant (MUT) clone in comparison to the wildtype (WT) clone. Elevated levels of glycolysis in response to hypoxia is notable in both WT and MUT clones. Data represents mean of three independent experiments ± SEM. Two-way ANNOVA with Fisher's LSD test was performed for statistical significance evaluations. * p value < 0.05. D) The basal OCR level was higher in the MUT clone in comparison to the WT clone and the difference was fully abolished following treatment of cells with an inhibitor of ATP synthase (oligomycin). E) The basal OCR in the WT clone was higher in comparison to the MUT clone under hypoxic conditions. Similar to normoxia, the difference was diminished following the inhibition of ATP synthase with oligomycin. Data shown was taken from one representative experiment of three independent experiments (biological replicates) performed. F) and G) Basal ECAR and PER levels were higher in PRSS23^{P230A} expressing cells compared to PRSS23^{WT} expressing cells which was reversed under hypoxic conditions showing higher levels in WT clone. Following inhibition of the respiratory chain, all clones regardless of expressing the wild-type or P230A variant of PRSS23 showed an increased PER and ECAR indicating elevated compensatory glycolysis. Data is from one representative experiment of three independent experiments (biological replicates) performed.

While OCR is directly indicative of mitochondrial oxidative respiration, production of lactate during glycolysis and production of CO_2 during respiration together contribute to extracellular acidification. [134] As mentioned earlier, the glycolytic rate and the proton efflux rate (PER) measured by the extracellular acidification rate (ECAR) are increased in PRSS23^{P230A} compared to PRSS23^{WT} expressing CaMCaA cells under normoxia. In hypoxic conditions, the glycolytic ATP production rate is significantly increased in PRSS23^{WT} expressing CaMCaA cells, which may compensate for the loss of energy and therefore, may explain the increased PER/ECAR compared to PRSS23^{P230A} expressing CaMCaA cells (Figures 14F and 14G).

Further analysis revealed that the decreased basal glycolytic ATP production rate observed in PRSS23^{WT} expressing CaMCaA cells was associated with elevated AMPK phosphorylation in these cells (refer to 3.8; Figure 11). These findings may endorse the proliferative advantage of PRSS23^{P230A} CaMCaA cells *in vitro* and the outgrowth of PRSS23^{P230A} tumors *in vivo*.

3.12 Electron microscopy reveals lower number of damaged mitochondria in PRSS23^{P230A} expressing CaMCaA cells

As PRSS23 represents a mitochondrial protein, which affects the mitochondrial ATP production, the morphology of the mitochondria was analyzed in PRSS23^{WT} and PRSS23^{P230A}

expressing CaMCaA cells using electron microscopy (Figure 15A). While all CaMCaA clones showed some level of cloudy swelling of mitochondria, the extent of this phenomenon and the number of severely damaged mitochondria were noticeably lower in PRSS23^{P230A} CaMCaA cells compared to PRSS23^{WT} control cells (Figure 15B). This observation could explain the increased mitochondrial ATP production rate in PRSS23^{P230A} expressing CaMCaA cells (refer to 3.11; Figure 14B). Notably, during visual evaluation of electron microscopic images of the CaMCaA clones, I realized the exclusive presence of vesicular structures in PRSS23^{WT} CaMCaA cells which may be correlated to the upregulation of lysosomal associated intracellular trafficking observed in expression profiling of these wildtype clones. (Figure 15C).

3.13 Identification of proteins interacting with PRSS23 variants

PRSS23 is a not-well studied member of the trypsin family of serine proteases and available analyses on its cellular functions are sparse and insufficient to understand the molecular mechanism through which PRSS23^{P230A} drives HCC development. Mass spectrometry was used to identify proteins differentially binding to either the PRSS23^{WT} or the PRSS23^{P230A} protein in CaMCaA cells (Figure 16A and 16B). In total, 1082 proteins were detected from which 57 solely were seen in IP samples from PRSS23^{WT} expressing CaMCaA cells and 86 were exclusively interacting with the variant PRSS23^{P230A}. 174 protein candidates were identified as differential binding partners of either WT or P230A variants (Figure 16C).

Using the STRING database [135], I performed pathway enrichment analysis of the interacting proteins which revealed that wild-type PRSS23 mainly interacts with components of intracellular and membrane trafficking. In line with this, several PRSS23^{WT} interacting proteins were associated with clathrin-coated endocytosis including AP2M1, HIP1R, AP1B1, EPN4, CLTA, PICALM, and STON2. These are involved in various cellular processes such as lysosomal degradation and internalization-mediated regulation of membrane receptors (Figure 16D and Supplementary Figure 1). On the other hand, the interacting partners with higher affinity for PRSS23^{P230A} were enriched for cell cycle and mRNA-associated cellular processes and thus mirror the data derived from expression profiling of PRSS23^{P230A} CaMCaA cells (Figure 16E and Supplementary Figure 2).
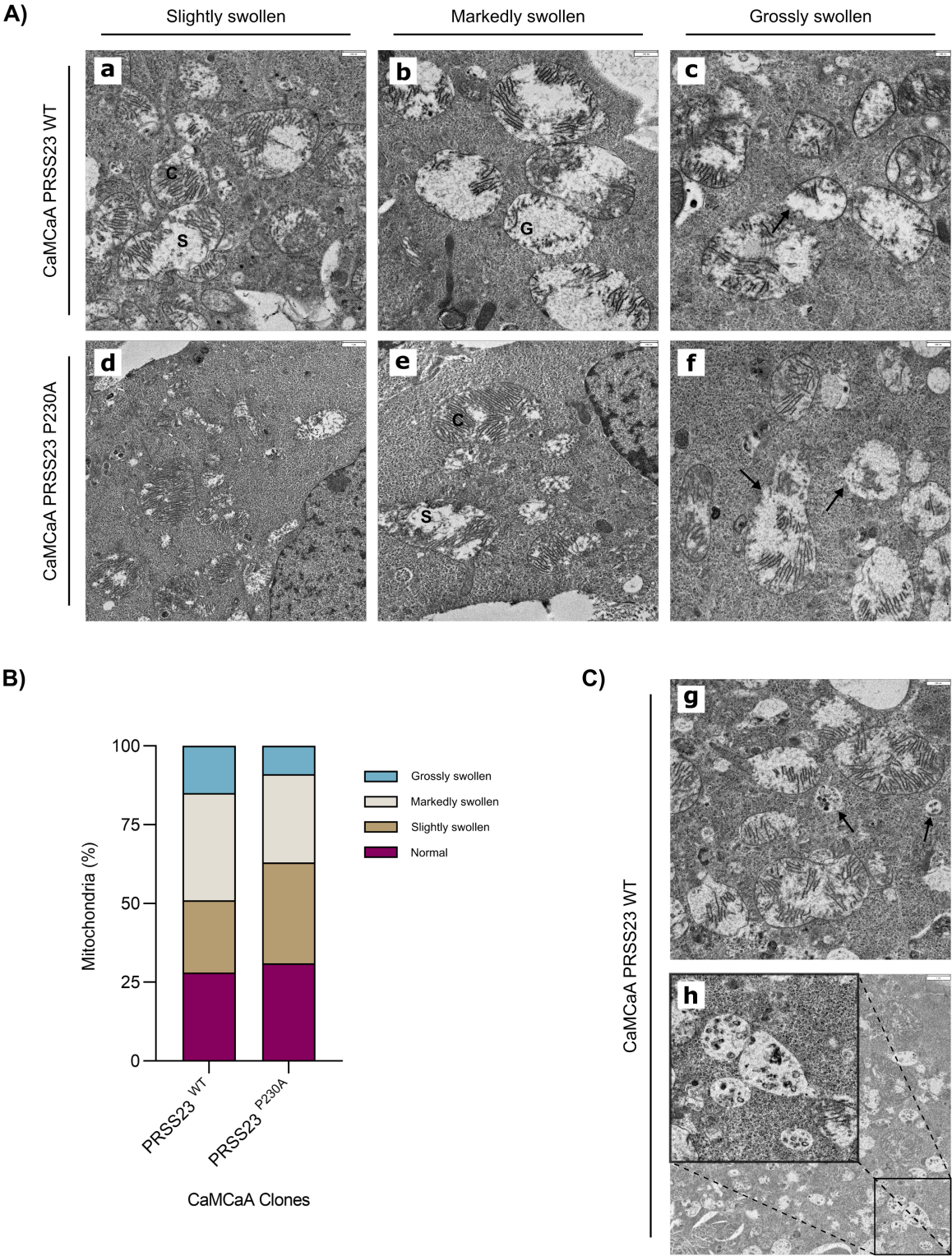
Figure 15. Electron microscopy of CaMCaA cells expressing PRSS23 variants. Electron microscopy of CaMCaA PRSS23[WT] and PRSS23[P230A] expressing clones was performed in close collaboration with Dr. Ingrid Haußer-Siller (Institute of Pathology Heidelberg). A) Electron microscopy images of mitochondria showing various extent of cloudy swelling (S). Slightly swollen mitochondria are characterized by peripherally placed cristae (C) and absence of intra-mitochondrial dense granules (G). Markedly swollen mitochondria feature

peripherally placed and disoriented cristae, patchy appearance. Dense granules (G) are present in abundance. Grossly swollen mitochondria represent loss of cristae and ruptured mitochondria membrane (arrow). [136] B) Electron microscopy images were evaluated for the number of normal and swollen mitochondria in each CaMCaA clone according to the morphological characterization by *Ultrastructural pathology of the cell and matrix* (Feroze N. Ghadially, 4th ed.). [136] 52 and 42 random microscopic images with the scale of 200-500 nm were evaluated for PRSS23^{WT} and PRSS23^{P230A} expressing cells, respectively. C) Vesicular structures were observed in remarkable numbers exclusively in CaMCaA PRSS23^{WT} cells. Scale bars, a) 500nm; b) 500nm; c) 200nm; d) 1µm; e) 500nm; f) 500nm; g) 500nm; h) 1µm.

Co-immunoprecipitation (Co-IP) analysis validated the exclusive interaction of Peripherin (PRPH) with C-terminally FLAG-tagged PRSS23^{P230A} (Figure 17E). The binding of DNA replication licensing factor MCM3 (MCM3) and GMP synthase (GMPS) (Figure 17B), and anaphase-promoting complex subunit 1 (ANAPC1) to C-terminally FLAG-tagged PRSS23^{WT} and PRSS23^{P230A} was detected in Western blot analysis of IP samples (Figure 17F). However, these results only partially matched with the findings from mass spectrometry. According to the mass spectrometry data, MCM3 was detected in more abundance in IP samples of C-terminally FLAG-tagged PRSS23^{P230A} (refer to Supplementary Table 1). However, the Western blot analysis did not show comparable differences in protein levels of MCM3 immunoprecipitated with C-terminally FLAG-tagged PRSS23^{WT} or PRSS23^{P230A}. Similarly, GMPS and ANAPC1 were detected by mass spectrometry exclusively in IP samples of C-terminally FLAG-tagged PRSS23^{P230A} and were absent in C-terminally FLAG-tagged PRSS23^{WT}. In contrast, similar levels of these proteins were detected in both C-terminally FLAG-tagged PRSS23^{P230A} and PRSS23^{WT} immunoprecipitations by Western blot. Moreover, the exclusive interaction of C-terminally FLAG-tagged PRSS23^{WT} with phospholipase D1 (PLD1) (Figure 17A) and C-terminally FLAG-tagged PRSS23^{P230A} with elastin microfibril interfacer 1 (EMILIN-1) (Figure 17C), DNA replication licensing factor MCM6 (MCM6) (Figure 13D), and O-linked N-acetylglucosaminyl transferase 110 kDa subunit (OGT) (Figure 17F) identified by the mass spectrometry could not be reproduced by the Western blot analysis of IP samples.
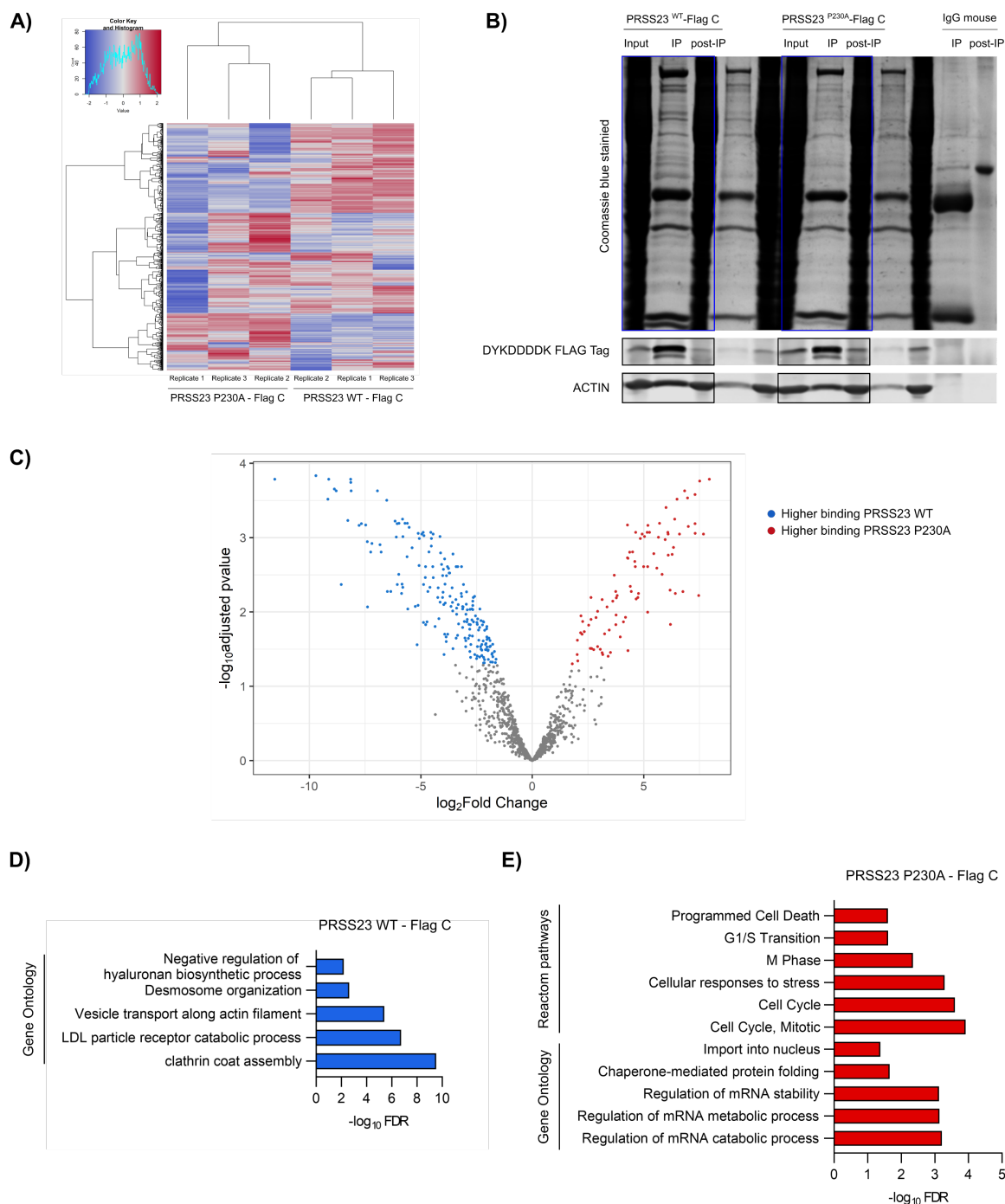
Figure 16. Substitution of proline at position 230 by alanine in PRSS23 alters the repertoire of interacting proteins. C-terminally FLAG-tagged PRSS23 was immunoprecipitated from whole cell lysates of CaMCaA wild-type (WT) and mutant (P230A) clones. Co-immunoprecipitated proteins were identified by mass spectrometry. A) Heatmap plot shows the clustering of three independent biological replicates of IP samples of each clone and the variation among wild-type and mutant samples. This Heatmap plot is kindly made by Martin Schneider (Genomics and Proteomics Core Facility of DKFZ, Heidelberg). B) Validation of immunoprecipitation of C-terminally FLAG-tagged PRSS23 variants by Coomassie Blue staining of SDS-Polyacrylamide gel and detection of FLAG tag via Western blot analysis. C) The volcano plot represents proteins differentially interacting with C-terminally FLAG-tagged PRSS23[WT] and PRSS23[P230A]. The threshold was set as logFC>2 or logFC<-2 and adjusted p-value <0.05. The blue dots indicate proteins with higher binding affinity to PRSS23[WT], while the red dots depict proteins with higher binding affinity to PRSS23[P230A]. D and E) Pathway enrichment analysis was

performed on proteins with higher binding affinity either to the C-terminally FLAG-tagged PRSS23^{WT} or the C-terminally FLAG-tagged PRSS23^{P230A} proteins. Top pathways with lowest FDR and highest enrichment score are listed.
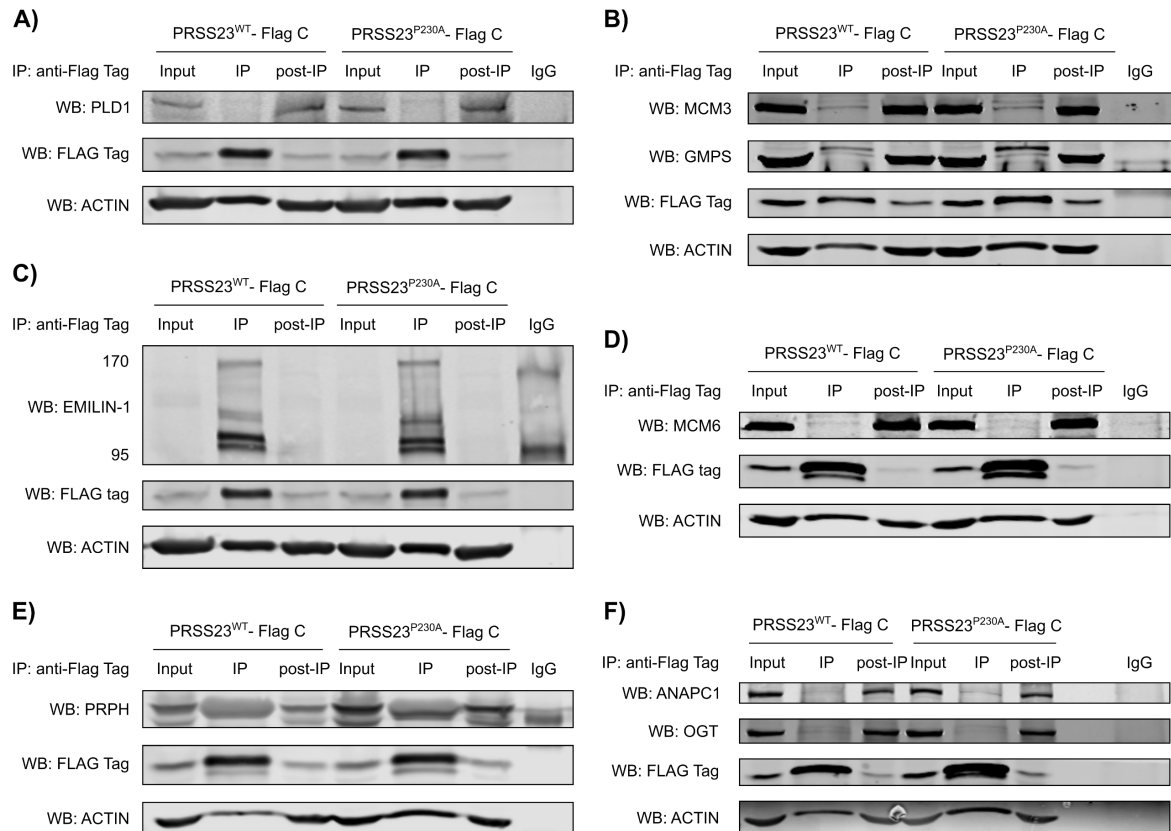
Figure 17. Co-immunoprecipitation of top ranked interacting partners with C-terminally FLAG-tagged PRSS23 wild-type or P230A protein. C-terminally FLAG-tagged PRSS23^{WT} or PRSS23^{P230A} were immunoprecipitated using a mouse anti-FLAG tag antibody and Western blot was used to detect the co-immmunoprecipitation (Co-IP) of top interacting candidates of mass spectrometry analysis. Interaction of A) PLD1 B) MCM3 and GMPS C) EMILIN-1 D) MCM6 E) PRPH F) ANAPC1 and OGT with C-terminally FLAG-tagged PRSS23^{WT} and PRSS23^{P230A} depicted by Western blots. Immunoprecipitation of C-terminally FLAG-tagged PRSS23$^{WT/P230A}$ was validated in all the blots using a rabbit anti-FLAG tag antibody. ACTIN was detected as loading control. Mouse IgG was used as isotype control.

3.14 Expression of PRSS23^{P230A} leads to elevated MYC activation in CaMCaA cells

As detailed above (section 3.6), expression profiling of PRSS23 wild-type and P230A mutant CaMCaA clones revealed the activation of Hippo and TGF-β signaling pathways upstream and cellular processes including cell cycle, DNA replication, RNA transcription, and ribosomal biogenesis to name a few of them, downstream of MYC signaling in PRSS23^{P230A} expressing CaMCaA cells connotating increased MYC activation in these mutant clones. The

observation that interacting partners of PRSS23 were also enriched in these signaling pathways (Figure 18A) further strengthened the hypothesis that MYC activation is fostered in PRSS23^{P230A} expressing CaMCaA cells. Therefore, I quantified the activation level of MYC in PRSS23^{WT} and PRSS23^{P230A} expressing CaMCaA clones using the TransAM$^®$ c-Myc kit (Active Motif), which detects the nuclear active form of MYC when bound to its target DNA. Strikingly, CaMCaA cells expressing PRSS23^{P230A} showed a markedly increased level of MYC activation compared to PRSS23^{WT} expressing CaMCaA cells (Figure 18D). Moreover, the protein level of nuclear MYC was significantly higher in PRSS23^{P230A} CaMCaA cells (Figure 18C).
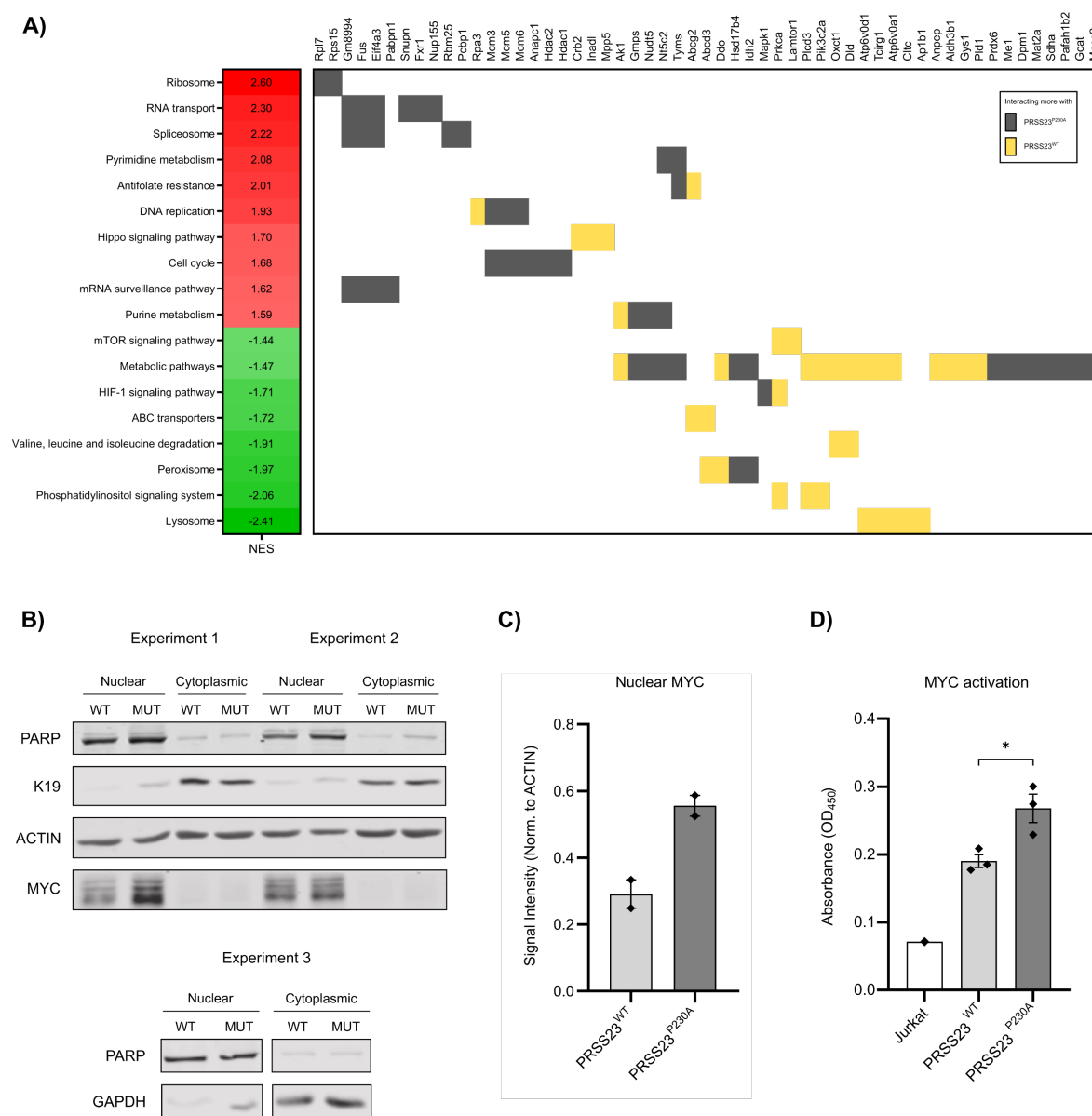
Figure 18. PRSS23^{P230A} fosters MYC activation in CaMCaA cells. A) Overlap of the differential interacting partners of the wild-type and the P230A variant of PRSS23 (as detected by mass spectrometry analysis) with

MYC associated signaling pathways reveals a differential regulation in gene expression between PRSS23^{WT} and PRSS23^{P230A} expressing CaMCaA clones. The KEGG's pathways are listed according to their NES value (left side). NES>0 represents enrichment in PRSS23^{P230A} CaMCaA cells and NES<0 represents enrichment in PRSS23^{WT} CaMCaA cells. NES, Normalized Enrichment Score. B) Western blot analysis of nuclear and cytoplasmic extracts of PRSS23^{WT} and PRSS23^{P230A} CaMCaA cells for poly (ADP-ribose) polymerase (PARP) as nuclear marker and keratin 19 (K19) and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) as cytoplasmic markers as well as MYC protein levels. ACTIN was detected as loading control. C) Densitometric analysis of nuclear MYC protein levels in CaMCaA PRSS23^{WT} and PRSS23^{P230A} clones. Data represent the signal intensity of MYC observed in experiments 1 and 2 (refer to panel B). Norm., Normalized. D) Nuclear extracts of unstimulated PRSS23^{WT} and PRSS23^{P230A} CaMCaA cells (Jurkat cells served as positive control) were analyzed by TransAM c-Myc assay (Active Motif). The level of MYC activation was significantly elevated in the PRSS23^{P230A} mutant expressing clone compared to PRSS23^{WT} expressing cells. Data represent mean ± SEM of three independent experiments. * p< 0.05, unpaired two-sided student's t-test.

4 DISCUSSION

Hepatocellular carcinoma is the most prevalent primary liver cancer and despite all the available therapeutic options one of the four major causes of cancer-related death worldwide. [1] One of the main limiting factors in treatment is diagnosis at late-stage of the disease, when curative surgery is no longer feasible. The other factor limiting the scope and efficacy of treatment alternatives of HCC is the presence of liver cirrhosis. This is present in 80% of patients and excludes both larger hepatic resections as well as many systemic therapy options of affected patients. [1, 21] Indeed, the most efficient therapies are applied at early disease stages. [22, 23] Therefore, diagnosis and treatment of HCC at early stages of the disease or even at premalignant stages allowing for prevention of malignant transformations has paramount importance.

The clonal evolution of cancer is morphologically represented by the step-wise process of human hepatocarcinogenesis. Therefore, identification of driver mutations arising in the cirrhotic liver and understanding the mechanism through which they foster the process of malignant transformation into early HCC, may potentially open new avenues for early diagnosis and treatment.

In this study, integration of whole exome sequencing and *in vivo* RNAi screening led to the identification of clonally expanded mutations during human hepatocarcinogenesis, which may drive liver cancer development and progression. In particular, the data generated within this thesis revealed a tumor suppressive role of PRSS23 in the context of activated MYC and AKT1 signaling. Further *in vivo* validations confirmed the oncogenic potential of the P230A variant of PRSS23 in MYC overexpressing *Trp53* heterozygous mice. *In vitro* investigation of the molecular mechanism underlying the tumor promoting potential of PRSS23[P230A] revealed that the enhanced activation of MYC and consequently activation of its downstream signaling pathways resulted in higher proliferation and survival capacity as well as improved cellular energetics. The results of this thesis imply that the loss of the tumor suppressive capabilities and eventually an additional gain of tumor promoting functions of PRSS23 is mediated by the P230A mutation. Functionally, this results in activation of MYC activity and boosters its tumor promoting effects during hepatocarcinogenesis.

4.1 Integration of cancer genomics and in vivo RNAi screening identified P230A variant of PRSS23 as driver mutation of hepatocarcinogenesis

The combination of comprehensive genetic profiling and functional genomics allowed for the identification of clonally expanded driver mutations in human hepatocarcinogenesis. In particular, the analysis revealed that mutant PRSS23 accelerates development in context of activated MYC and AKT1 signaling. In addition, the findings of the present thesis suggest a potential tumor suppressive function of VWA5A, CLMP, and DLK1 during HCC development. Although not focused in this thesis, there is consistency with the published data proposing a tumor suppressive role for VWA5A in melanoma [137] and nasopharyngeal carcinoma [138]. However, the data regarding the role of Dlk1 are heterogeneous. Whereas several studies suggested oncogenic functions of DLK1 in hepatocellular carcinoma [139-141] and other cancer types [142, 143], Kawakami et al. and the present study highlighted a putative tumor suppressive role of DLK1, both in human renal cell carcinoma [144] and hepatocellular carcinoma, respectively. In addition, this thesis provides the first experimental data on potential tumor suppressive functions of CLMP and PRSS23 in HCC. While previous studies have associated high PRSS23 expression levels with tumorigenicity in several cancer types [109-114], the precise function of both proteins has not been deciphered yet.

Validation experiments, in which each individual variant detected in the human HCC nodule was expressed *in vivo* in the mosaic mouse model confirmed the tumorigenic potential of the P230A variant of PRSS23 in MYC-induced HCC. Furthermore, the present data suggest a tumor suppressive function of wild-type PRSS23 and demonstrate that a loss of its regular function due to the P230A mutation promotes tumor development. Although the present results did not reveal an indication that the *Vwa5a*, *Clmp*, and *Dlk1* act as tumor suppressor genes during HCC development, the validation experiments were confined to the MYC overexpressing HCC model used for validation purposes. Future research may consider other relevant genetic backgrounds (e.g., YAP, TP53) as well, which may allow to extend the current findings and provide a more comprehensive molecular picture.

It is interesting that MYC overexpression HCC was the only genetic background in mice supporting the tumorigenic effects of the P230A variant of PRSS23. This could imply that the P230A mutation of PRSS23 may act as a secondary event fostering MYC activation to a level required for malignant transformation of murine hepatocytes. [145] In line with this data, MYC overexpression is frequently observed in human HCC. [63]

Within a collaboration initiated during this study a computational 3D model of PRSS23 was generated. In this model, proline 230 is located sufficiently close to the aspartic acid residue (D246) of the catalytic triad and its substitution with alanine may result in establishment of a salt bridge between A230 and D246. In serine proteases, the nucleophilicity of the catalytic serine strongly relies on the Asp-His-Ser catalytic triad particularly on the negative charge of the aspartic acid residue known as the charge relay system. [87] Therefore, establishment of the latter salt bridge between A230 and D246 in PRSS23 may disturb the charge relay system of the catalytic triad and as a result, impairing the nucleophilicity of the catalytic serine residue. This may impede the catalytic activity of the enzyme. Salt bridges are essential for establishment of the protein structure and function and are considered to be involved in protein stability and target recognition. [146] Therefore, it seems likely that the P230A amino acid substitution may have a stabilizing effect on PRSS23[P230A] folding. In addition, it seems reasonable that binding of substrates and/or interacting proteins may be affected. Thus, the 3D model may structurally explain the loss of the tumor suppressive function of PRSS23[P230A] and its altered binding affinity to interacting proteins compared to the wild-type protein seen in the mass spectrometry experiments following PRSS23 immunoprecipitation (refer to 3.3 and 3.13). Biochemically, the formation of a salt bridge between A230 and D246 is not probable due to neutral charge of alanine at intracellular pH 7.0. [147, 148] However, the overall charge of alanine at acidic pH is positive suggesting that the formation of this salt bridge in PRSS23[P230A] may be pH-dependent. This highlights that (un)favored effects of the P230A mutation on PRSS23 may be confined to acidic organelles or milieus such as lysosomes and secretory vesicles or at a greater scale, to the hypoxic microenvironment of growing tumor nodules. [149] One of the most relevant works supporting this hypothesis is the one published by Kokkonen et al. showing that altered Golgi pH may result in mistargeting of membrane proteins to the basolateral surface of cells and loss of functional polarity of plasma membrane in colorectal cancer cells. [150]

One limitation of the present study was that the effect of the P230A mutation on proteolytic activity of PRSS23 was not analyzed. Lack of a well-characterized substrate for PRSS23 precluded this focus. Nevertheless, the mass spectrometry-based analysis identified a set of differentially interacting partners between the P230A and wildtype variants of PRSS23 and thus provided evidence that the spectrum of proteins as well as the binding affinity of particular partners for PRSS23 is dependent on the nucleotide sequence of the PRSS23 protein. However, another interpretation of these findings may be that the P230A mutation may not

only have impact on the protein structure but also the potentially altered proteolytic function of PRSS23 may modify the proteomics composition of cells expressing PRSS23[P230A] compared to cells expressing PRSS23[WT] and that the different spectrum of interacting proteins between PRSS23[WT] and PRSS23[P230A] expressing CaMCaA cells may be a projection of this phenomenon.

4.2 PRSS23[P230A] confers a selective growth advantage to CaMCaA cells

The second focus of this research project was to understand the molecular mechanism by which PRSS23[P230A] promotes hepatocellular carcinoma. Functional analysis of CaMCaA cells expressing PRSS23[P230A] provided evidence on the *in vivo* tumor promoting potential of PRSS23[P230A] previously reported in this study. In particular, PRSS23[P230A] increased the proliferation and cell viability of expressing CaMCaA cells. This phenotype may be explained by the higher fraction of cells in S and G2/M phases of the cell cycle suggesting that the inhibition of G1/S checkpoint may be enhanced in PRSS23[P230A] compared to PRSS23[WT] expressing cells. This idea is further corroborated by the finding that PRSS23[P230A] expressing CaMCaA cells showed elevated total ATP production rate compared to their wild-type counterparts and thus might have a higher ATP reservoir. Interestingly, high intracellular ATP level has inhibitory effect on AMPK activation. [151, 152] In line with this, reduced levels of AMPK phosphorylation were observed in PRSS23[P230A] compared to PRSS23[WT] expressing cells. This implies an increased blockade of the G1/S checkpoint in PRSS23[P230A] cells. On the contrary, it seems reasonable to speculate that in the context of a shortfall of nutrients and a high tumor cell density, the tumor suppressive function of PRSS23[WT] may prevent cell cycle progression. [153, 154] However, this analysis only provides a portrayal of the cell cycle distribution in the population of cells at 72 hours of culture with the lack of information on cell cycle kinetics over a broader period of time. [155] Therefore, further investigation of cell cycle markers throughout all the proliferative phases will better explain the current findings.

Invasion and metastasis are well-known hallmarks of cancer. [156] In this study, the findings of the cell invasion assay suggested that PRSS23[P230A] may also promote tumorigenesis via conferring invasiveness to premalignant hepatocytes. Of note, PRSS23 is a member of the S1A family of serine proteases. These are primarily considered to be involved in processing of extracellular matrix (ECM) proteins. [87, 88] In this study, the concentration of secreted PRSS23 was not different between wild-type and mutant clones *in vitro*. Thus, the elevated invasiveness observed in CaMCaA PRSS23[P230A] cells is not due to increased levels of secreted

PRSS23[P230A]. However, it remains unclear whether the P230A variant may also have a gain of function, which allows PRSS23 to modulate the ECM in favor of tumor cell invasion or whether the loss of the catalytic function may support the invasiveness of transformed hepatocytes through unknown intracellular mechanisms. One main concern about these findings is that they cannot simply be translated into *in vivo* situations. In physiological conditions, intracellular signaling together with the composition of the ECM regulates the secretion of ECM remodeling factors such as proteases. [157] In this respect, conventional isogenic monolayer cell cultures fail to resemble the pathophysiological condition *in vivo*. To at least improve the system in this sense, application of collagen or Matrigel® matrices as an alternative culture system may be beneficial to study the secreted PRSS23 levels of CaMCaA cells and might be able to provide more insights into these processes.

There is not much evidence on the intracellular localization of PRSS23. Nonetheless, there were reports that PRSS23 may be a nuclear protein [106], but it has also been described in the endoplasmic reticulum lumen [158, 159], and extracellular exosomes [160]. The results of this study cast a new light on the cellular localization of PRSS23. In particular, PRSS23 was found enriched in the mitochondrial fraction of CaMCaA cells compared to the cytoplasmic and nuclear compartments. However, these findings suffer from the same limitations associated with the lack of antibodies reliably detecting PRSS23.

4.3 Expression of PRSS23[P230A] benefits cellular energetics via enhancing ATP production rate in CaMCaA cells

The current work showed elevated ATP production in CaMCaA cells expressing PRSS23[P230A] compared to PRSS23[WT] expressing control cells. This increase was mainly due to the upregulation of mitochondrial oxidative phosphorylation as well as glycolysis in PRSS23 mutant cells. Another finding was that the inhibition of complex V (ATP synthase) of the mitochondrial respiratory chain completely abolished the difference in the cellular respiration rate (measured as OCR) between wild-type and mutant clones. From these results it became evident that PRSS23 acts as an upstream regulator of mitochondrial respiration and that the elevated ATP levels in CaMCaA PRSS23[P230A] cells are most likely linked to boosted ATP synthase activity and not related to the complexes I and III of the mitochondrial electron transfer chain (ETC).

This study further demonstrated that this metabolic advantage of PRSS23[P230A] cells was reversed under hypoxia, implying to higher sensitivity of PRSS23[P230A] clones to hypoxic

conditions and that the proliferative advantage of mutant compared to wild-type clones may be related to the higher rate of oxidative phosphorylation. One explanation for this phenotype is provided by the work of Robinson et al. (2012), which proposed that elevated oxidative phosphorylation supports cell viability of glucose-starved cancer cells via higher supply of ATP. Additionally, they demonstrated that a cloudy swelling was present in most of the mitochondria in glucose-maintained cancer cells. In addition, glucose-starved cancer cells possessed a greater number of mitochondria and the morphology of mitochondria showed densely and regularly packed cristae, a known sign of their metabolic dependency on oxidative phosphorylation. [161] The later observation ties well with our electron microscopy findings of this thesis, which revealed that the number of mitochondria showing cloudy swelling was significantly higher in PRSS23^{WT}- compared to PRSS23^{P230A}- expressing CaMCaA cells, while the mutant clones had greater number of mitochondria with normal cristae morphology. Taken together, the present thesis provides evidence that PRSS23^{WT} expressing CaMCaA control cells, like many other tumor-derived cell lines, activate the AMPK signaling. [162] In addition, deprivation of nutrients and/or hypoxia result in metabolic reprogramming towards enhanced glycolysis in PRSS23^{WT} cells as a compensatory mechanism to depleted ATP supply. In contrast, PRSS23^{P230A} expressing CaMCaA cells retain the ability to metabolically switch to oxidative phosphorylation. Altogether, these findings endorsed by the gene expression profiling of CaMCaA clones denote that the metabolism of PRSS23^{WT} expressing CaMCaA cells is mainly dependent on glycolysis, while expression of PRSS23^{P230A} induce a remodeling of mitochondrial cristae fostering the capacity of oxidative phosphorylation, which may explain the higher proliferation capacity of PRSS23^{P230A} cells *in vitro* and probably also the clonal expansion of PRSS23^{P230A} tumor cells *in vivo*. However, hypoxia leads to acidification of the milieu, which may – as outlined above – affect the structure of the catalytic triad and thus may be able to revert the energetic advantage observed for PRSS23^{P230A} expressing HCC cells under normoxic conditions.

4.4 PRSS23^{P230A} contributes to tumor growth through a mechanism involving MYC activation

Expression profiling of CaMCaA cells expressing either the wild-type or the mutant P230A variant of PRSS23 revealed differential regulation of signaling pathways and cellular processes. The top ranked up- and down-regulated pathways in CaMCaA PRSS23^{P230A} clones were ribosome and ribosomal biogenesis vs. lysosomal-associated transport and catabolism and autophagy, respectively. This altered regulatory pattern implicated that the augmented

activity of MYC in PRSS23[P230A] expressing CaMCaA cells may play a role. The oncoprotein MYC is well-recognized as the key regulator of ribosomal biogenesis and protein translation and its overexpression with a consequent aberrant activation is strongly associated with cancer. [163] Furthermore, suppression of lysosomal and autophagic genes is known to be downstream of MYC signaling and negatively regulated by histone deacetylase (HDAC)/MYC and the microphthalmia (MiT/TFE) transcription factors. [164] Nuclear localization and activity of MiT/TFE transcription factors is tightly regulated by mTORC1 complex in a nutrient-dependent manner. [165] This is consistent with elevated levels of pAMPK and upregulation of lysosomal and autophagic pathways in CaMCaA PRSS23[WT] cells (See Figures 11 and 9). In addition, these evidences further supported the hypothesis of elevated MYC activity in PRSS23[P230A] expressing CaMCaA clones. A further novel finding of this thesis is the higher levels of active MYC in nuclear protein extracts of PRSS23[P230A] expressing CaMCaA cells compared to cells expressing PRSS23[WT]. Importantly, HDAC1 and HDAC2 were recognized as exclusive interacting partners of PRSS23[P230A] in mass spectrometry analysis suggesting that PRSS23[P230A] may induce elevated MYC activation in CaMCaA cells via modulating HDACs.

4.5   Conclusions and future perspectives

Taken together, the present thesis identified and validated clonally expanded driver gene mutations of human hepatocarcinogenesis via integration of whole exome sequencing and *in vivo* RNA interference screening. The single gene validation *in vivo* confirmed that the P230A variant of PRSS23 acts as driver of malignant transformation of hepatocytes in the context of activated MYC signaling. Further *in vitro* functional and biochemical analysis of PRSS23[P230A] expressing CaMCaA cells illustrated that PRSS23[P230A] is able to switch of cellular energetics towards oxidative phosphorylation providing CaMCaA cells with a higher ATP supply. This very likely promotes the proliferative capacity and cell viability of PRSS23[P230A] expressing HCC cells. In addition, the elevated ATP level in PRSS23[P230A] expressing cells lead to downregulation of AMPK signaling and consequently, reduce antagonistic effects of MiT/TFE transcription factors on MYC activation. The latter is reflected by an elevated ribosomal biogenesis and protein translation rate of PRSS23[P230A] expressing CaMCaA cells compared to elevated lysosomal degradation and autophagy in PRSS23[WT] expressing counterparts. However, with our current understanding it still remains elusive at which point of this signaling cascade PRSS23[P230A] interferes and fosters MYC signaling. Further validation of mass spectrometry analysis regarding PRSS23 interacting proteins is currently ongoing and will lead to identification of robust direct binding partners of PRSS23.

These new findings will hopefully help to decipher the exact mechanism by which PRSS23^{P230A} modulates the cross talk between the AMPK, PI3K-AKT-mTOR, and MYC signaling cascades, which obviously confers the growth advantage of PRSS23^{P230A} expressing HCC cells and leads to the clonal outgrowth of Dysplastic Nodules into hepatocellular carcinoma during hepatocarcinogenesis.

SUPPLEMENTARY DATA

Supplementary Table 1. List of interacting proteins of PRSS23 selected for further validation.

Gene	Protein Name	logFC	adj.P.Value	Contrast Max Score
Prph	Peripherin	6.61	0.0009	0.91
Emilin1	EMILIN-1	6.08	0.0008	0.65
Gmps	GMP synthase	4.40	0.0045	0.37
Mcm6	DNA replication licensing factor MCM6	3.95	0.0217	0.41
Ogt	OGT	3.09	0.1717	0.37
Mcm3	DNA replication licensing factor MCM3	1.68	0.0957	0.69
Anapc1	Anaphase-promoting complex subunit 1	1.25	0.3258	0.37
Pld1	Phospholipase D1	-9.13	0.0002	0.73

The logFC represents the differential intensity of the detected protein between PRSS23^{WT} and PRSS23^{P230A} IP samples. The Contrast Max Score is a quality value ranging from 1 to 0 (high to low trustworthiness). This score is developed in-house to represent the robustness of the mass spectrometry detection and merges several parameters including number of detected unique peptides, sequence coverage, abundance of the protein, identification type, etc. EMILIN-1, MCM6, OGT, and ANAPC1 were only detected in PRSS23^{P230A} IP samples. PLD1 was only detected in PRSS23^{WT} IP samples.

Abbreviations: logFC, log fold-change; adj.P.Value, adjusted p.value.
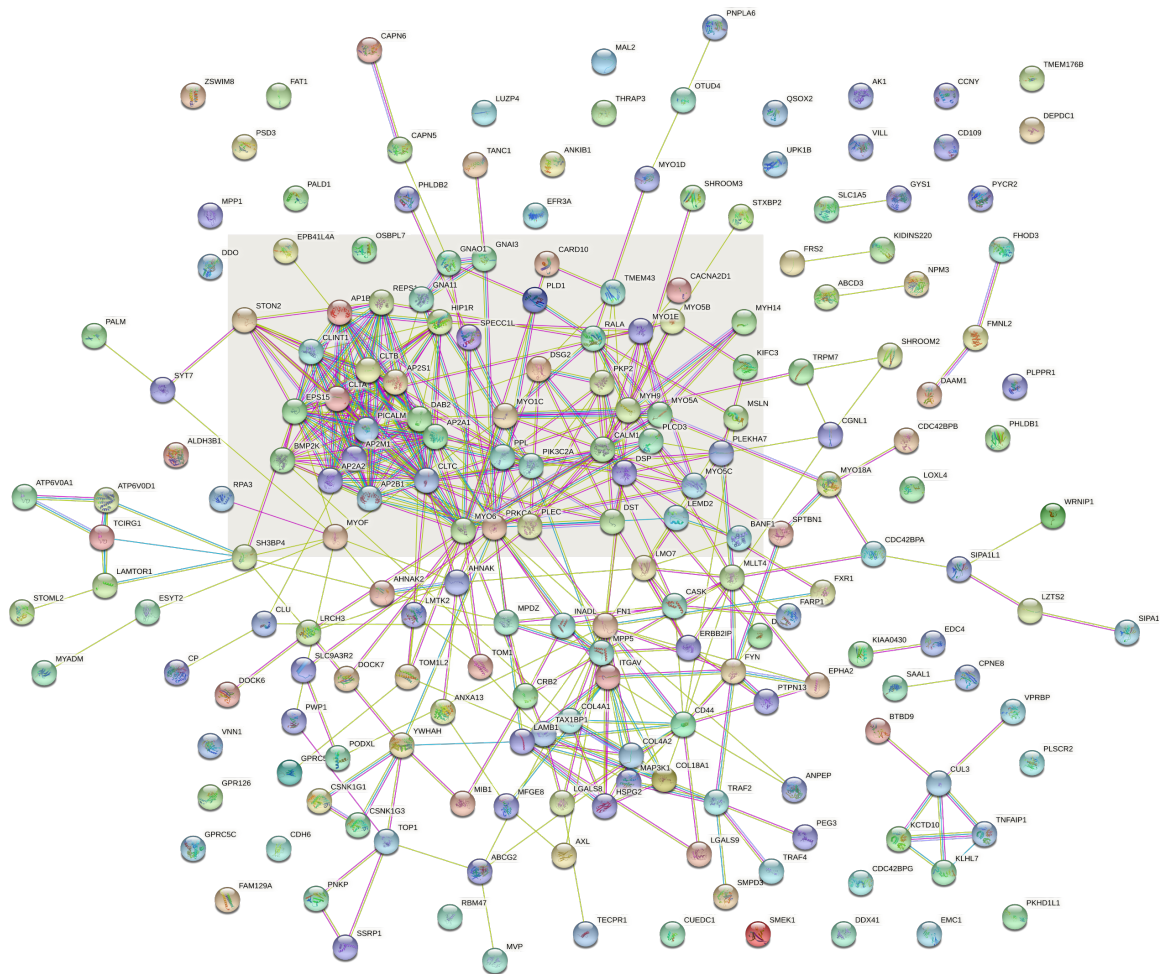
Supplementary Figure 1. STRING network analysis of the mass spectrometry detected proteins interacting more with wild-type PRSS23. The protein names were uploaded on the database and full STRING network analysis was performed. This analysis considers both functional and physical protein associations and was set to derive interactions from database knowledge, high-throughput lab experiments, automated textmining with the minimum required interaction score of 0.04 (medium confidence). [135] Distinct clustering of proteins associated with clathrin-coated vesicle-mediated endocytosis and intracellular trafficking is notable (with highlighted background).
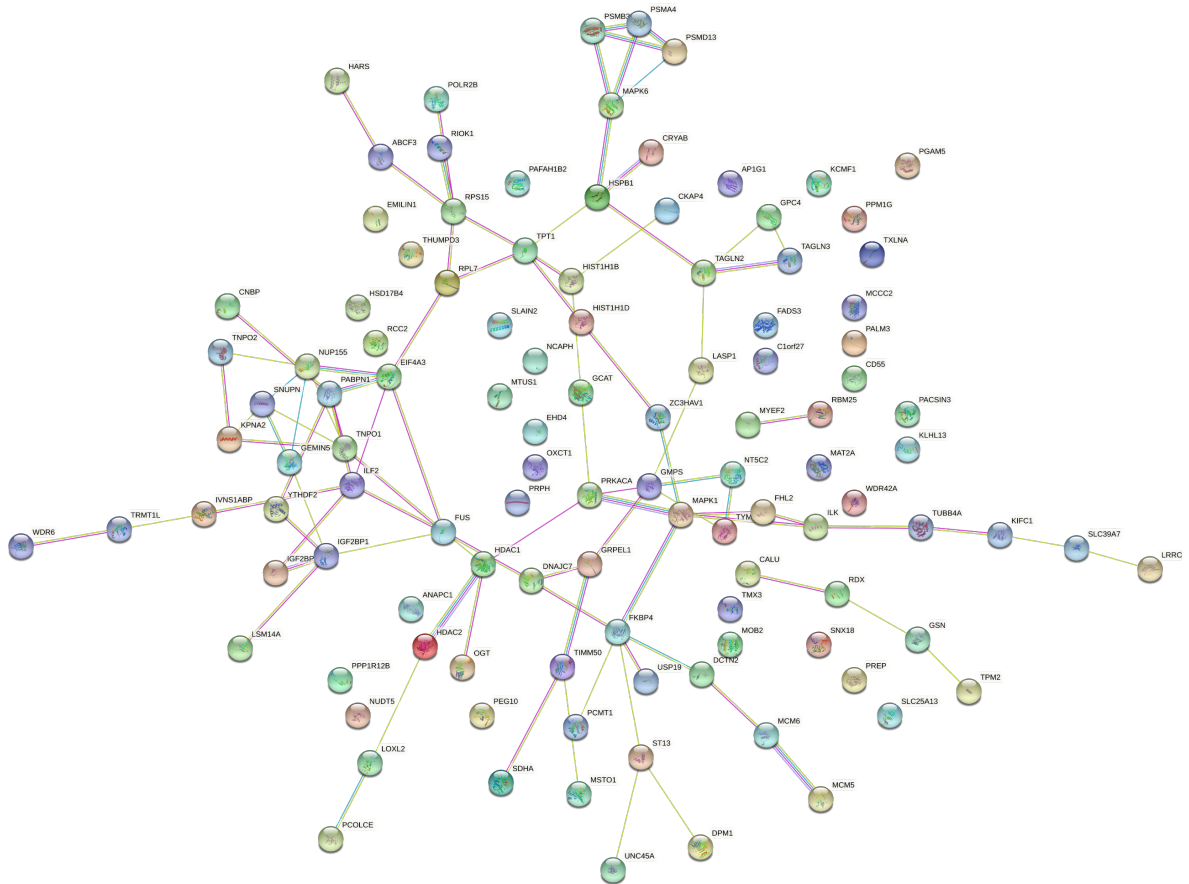
Supplementary Figure 2. STRING network analysis of the mass spectrometry detected proteins interacting more with P230A variant of PRSS23. The protein names were uploaded on the database and full STRING network analysis was performed. This analysis considers both functional and physical protein associations and was set to derive interactions from database knowledge, high-throughput lab experiments, automated textmining with the minimum required interaction score of 0.04 (medium confidence). [135] A few disperse clusters linked to proteasome complex, spliceosome, nuclear transport, and ribosomes are detectable.

SUPPLEMENTARY DATA

REFERENCES

1.   Llovet, J.M., et al., *Hepatocellular carcinoma.* Nat Rev Dis Primers, 2021. **7**(1): p. 6.

2.   Kew, M.C., *Synergistic interaction between aflatoxin B1 and hepatitis B virus in hepatocarcinogenesis.* Liver Int, 2003. **23**(6): p. 405-9.

3.   Ganne-Carrie, N. and P. Nahon, *Hepatocellular carcinoma in the setting of alcohol-related liver disease.* J Hepatol, 2019. **70**(2): p. 284-293.

4.   Younossi, Z.M., et al., *The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: A systematic review and meta-analysis.* J Hepatol, 2019. **71**(4): p. 793-801.

5.   Polyzos, S.A., J. Kountouras, and C.S. Mantzoros, *Obesity and nonalcoholic fatty liver disease: From pathophysiology to therapeutics.* Metabolism, 2019. **92**: p. 82-97.

6.   Younossi, Z.M., et al., *Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes.* Hepatology, 2016. **64**(1): p. 73-84.

7.   Huang, D.Q., H.B. El-Serag, and R. Loomba, *Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention.* Nat Rev Gastroenterol Hepatol, 2021. **18**(4): p. 223-238.

8.   Desai, A., et al., *Hepatocellular carcinoma in non-cirrhotic liver: A comprehensive review.* World J Hepatol, 2019. **11**(1): p. 1-18.

9.   Foerster, F., et al., *NAFLD-driven HCC: Safety and efficacy of current and emerging treatment options.* J Hepatol, 2021.

10.   Flemming, J.A., et al., *NAFLD and Alcohol-Associated Liver Disease Will Be Responsible for Almost All New Diagnoses of Cirrhosis in Canada by 2040.* Hepatology. **n/a**(n/a).

11.   Phipps, M., et al., *Gender Matters: Characteristics of Hepatocellular Carcinoma in Women From a Large, Multicenter Study in the United States.* Am J Gastroenterol, 2020. **115**(9): p. 1486-1495.

12.   Rich, N.E., et al., *Hepatocellular Carcinoma Incidence Is Decreasing Among Younger Adults in the United States.* Clin Gastroenterol Hepatol, 2020. **18**(1): p. 242-248.e5.

13.   Rich, N.E., et al., *Racial and Ethnic Differences in Presentation and Outcomes of Hepatocellular Carcinoma.* Clin Gastroenterol Hepatol, 2019. **17**(3): p. 551-559.e1.

14.   Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2020). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed [29 Sept 2021].

15.   Islami, F., et al., *Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics.* J Natl Cancer Inst, 2021.

16.   Liao, S.H., et al., *Long-term effectiveness of population-wide multifaceted interventions for hepatocellular carcinoma in Taiwan.* J Hepatol, 2021. **75**(1): p. 132-141.

17.   Kanwal, F., et al., *Risk of Hepatocellular Cancer in HCV Patients Treated With Direct-Acting Antiviral Agents.* Gastroenterology, 2017. **153**(4): p. 996-1005.e1.

18. Moussa, I., et al., *Dietary Patterns and Hepatocellular Carcinoma Risk among US Adults.* Nutrients, 2021. **13**(6).

19. Lange, N.F., P. Radu, and J.F. Dufour, *Prevention of NAFLD-associated HCC: Role of lifestyle and chemoprevention.* J Hepatol, 2021.

20. Seyda Seydel, G., et al., *Economic growth leads to increase of obesity and associated hepatocellular carcinoma in developing countries.* Ann Hepatol, 2016. **15**(5): p. 662-72.

21. Fattovich, G., et al., *Hepatocellular carcinoma in cirrhosis: Incidence and risk factors.* Gastroenterology, 2004. **127**(5, Supplement 1): p. S35-S50.

22. Llovet, J.M., C. Brú, and J. Bruix, *Prognosis of Hepatocellular Carcinoma: The BCLC Staging Classification.* Semin Liver Dis, 1999. **19**(03): p. 329-338.

23. Llovet, J.M., et al., *Trial Design and Endpoints in Hepatocellular Carcinoma: AASLD Consensus Conference.* Hepatology, 2021. **73**(S1): p. 158-191.

24. Casak, S.J., et al., *FDA Approval Summary: Atezolizumab Plus Bevacizumab for the Treatment of Patients with Advanced Unresectable or Metastatic Hepatocellular Carcinoma.* Clin Cancer Res, 2021. **27**(7): p. 1836-1841.

25. Finn, R.S., et al., *Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma.* N Engl J Med, 2020. **382**(20): p. 1894-1905.

26. Llovet, J.M., et al., *Sorafenib in advanced hepatocellular carcinoma.* N Engl J Med, 2008. **359**(4): p. 378-90.

27. Bruix, J., et al., *Prognostic factors and predictors of sorafenib benefit in patients with hepatocellular carcinoma: Analysis of two phase III studies.* J Hepatol, 2017. **67**(5): p. 999-1008.

28. Scott, L.J., *Lenvatinib: First Global Approval.* Drugs, 2015. **75**(5): p. 553-560.

29. Rimini, M., et al., *Lenvatinib versus Sorafenib as first-line treatment in hepatocellular carcinoma: a multi-institutional matched case-control study.* Hepatol Res, 2021.

30. Nair, A., et al., *FDA Supplemental Approval Summary: Lenvatinib for the Treatment of Unresectable Hepatocellular Carcinoma.* Oncologist, 2021. **26**(3): p. e484-e491.

31. Bruix, J., et al., *Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial.* Lancet, 2017. **389**(10064): p. 56-66.

32. Heo, Y.-A. and Y.Y. Syed, *Regorafenib: A Review in Hepatocellular Carcinoma.* Drugs, 2018. **78**(9): p. 951-958.

33. Abou-Alfa, G.K., et al., *Cabozantinib in Patients with Advanced and Progressing Hepatocellular Carcinoma.* N Engl J Med, 2018. **379**(1): p. 54-63.

34. Deeks, E.D., *Cabozantinib: A Review in Advanced Hepatocellular Carcinoma.* Targeted Oncology, 2019. **14**(1): p. 107-113.

35. Zhu, A.X., et al., *Ramucirumab after sorafenib in patients with advanced hepatocellular carcinoma and increased α-fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial.* Lancet Oncol, 2019. **20**(2): p. 282-296.

36. *EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma.* Journal of Hepatology, 2018. **69**(1): p. 182-236.

37. Marquardt, J.U., J.B. Andersen, and S.S. Thorgeirsson, *Functional and genetic deconstruction of the cellular origin in liver cancer.* Nature Reviews Cancer, 2015. **15**(11): p. 653-667.

38. *Terminology of nodular hepatocellular lesions.* Hepatology, 1995. **22**(3): p. 983-93.

39. Desmet, V.J., *East–West pathology agreement on precancerous liver lesions and early hepatocellular carcinoma.* Hepatology, 2009. **49**(2): p. 355-357.

40. Borzio, M., et al., *Impact of large regenerative, low grade and high grade dysplastic nodules in hepatocellular carcinoma development.* J Hepatol, 2003. **39**(2): p. 208-14.

41. Neoplasia, I.C.G.f.H., *Pathologic diagnosis of early hepatocellular carcinoma: A report of the international consensus group for hepatocellular neoplasia.* Hepatology, 2009. **49**(2): p. 658-664.

42. Kojiro, M., *'Nodule-in-nodule' appearance in hepatocellular carcinoma: its significance as a morphologic marker of dedifferentiation.* Intervirology, 2004. **47**(3-5): p. 179-83.

43. Longerich, T., *[Hepatocellular carcinoma].* Pathologe, 2020. **41**(5): p. 478-487.

44. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome.* Nature, 2009. **458**(7239): p. 719-24.

45. Torrecilla, S., et al., *Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma.* J Hepatol, 2017. **67**(6): p. 1222-1231.

46. Llovet, J.M., et al., *Hepatocellular carcinoma.* Nat Rev Dis Primers, 2016. **2**: p. 16018.

47. Zucman-Rossi, J., et al., *Genetic Landscape and Biomarkers of Hepatocellular Carcinoma.* Gastroenterology, 2015. **149**(5): p. 1226-1239.e4.

48. Chiang, D.Y., et al., *Focal gains of VEGFA and molecular classification of hepatocellular carcinoma.* Cancer Res, 2008. **68**(16): p. 6779-88.

49. Hoshida, Y., et al., *Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma.* Cancer Res, 2009. **69**(18): p. 7385-92.

50. Boyault, S., et al., *Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets.* Hepatology, 2007. **45**(1): p. 42-52.

51. Ally, A., et al., *Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma.* Cell, 2017. **169**(7): p. 1327-1341.e23.

52. Llovet, J.M., et al., *Molecular therapies and precision medicine for hepatocellular carcinoma.* Nat Rev Clin Oncol, 2018. **15**(10): p. 599-616.

53. Llovet, J.M., et al., *Advances in targeted therapies for hepatocellular carcinoma in the genomic era.* Nat Rev Clin Oncol, 2015. **12**(8): p. 436.

54. Lee, J.S., et al., *Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling.* Hepatology, 2004. **40**(3): p. 667-76.

55. Calderaro, J., et al., *Molecular and histological correlations in liver cancer.* J Hepatol, 2019. **71**(3): p. 616-630.

56.   Lachenmayer, A., et al., *Wnt-pathway activation in two molecular classes of hepatocellular carcinoma and experimental modulation by sorafenib.* Clin Cancer Res, 2012. **18**(18): p. 4997-5007.

57.   Ziol, M., et al., *Macrotrabecular-massive hepatocellular carcinoma: A distinctive histological subtype with clinical relevance.* Hepatology, 2018. **68**(1): p. 103-112.

58.   Rebouissou, S. and J.C. Nault, *Advances in molecular classification and precision oncology in hepatocellular carcinoma.* J Hepatol, 2020. **72**(2): p. 215-229.

59.   Erstad, D.J., B.C. Fuchs, and K.K. Tanabe, *Molecular signatures in hepatocellular carcinoma: A step toward rationally designed cancer therapy.* Cancer, 2018. **124**(15): p. 3084-3104.

60.   Schulze, K., et al., *Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets.* Nature Genetics, 2015. **47**(5): p. 505-511.

61.   Guichard, C., et al., *Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma.* Nat Genet, 2012. **44**(6): p. 694-8.

62.   Longerich, T., et al., *Oncogenetic tree modeling of human hepatocarcinogenesis.* Int J Cancer, 2012. **130**(3): p. 575-83.

63.   Schlaeger, C., et al., *Etiology-dependent molecular mechanisms in human hepatocarcinogenesis.* Hepatology, 2008. **47**(2): p. 511-20.

64.   Zender, L., et al., *Cancer gene discovery in hepatocellular carcinoma.* J Hepatol, 2010. **52**(6): p. 921-9.

65.   Gordon, J.W., et al., *Genetic transformation of mouse embryos by microinjection of purified DNA.* Proc Natl Acad Sci U S A, 1980. **77**(12): p. 7380-4.

66.   Newell, P., et al., *Experimental models of hepatocellular carcinoma.* J Hepatol, 2008. **48**(5): p. 858-79.

67.   Zender, L., et al., *An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer.* Cell, 2008. **135**(5): p. 852-64.

68.   Seehawer, M., et al., *Necroptosis microenvironment directs lineage commitment in liver cancer.* Nature, 2018. **562**(7725): p. 69-75.

69.   Pontes-Quero, S., et al., *Dual ifgMosaic: A Versatile Method for Multispectral and Combinatorial Mosaic Gene-Function Analysis.* Cell, 2017. **170**(4): p. 800-814.e18.

70.   Jonkers, J. and A. Berns, *Conditional mouse models of sporadic cancer.* Nat Rev Cancer, 2002. **2**(4): p. 251-65.

71.   Zender, L., et al., *Generation and analysis of genetically defined liver carcinomas derived from bipotential liver progenitors.* Cold Spring Harb Symp Quant Biol, 2005. **70**: p. 251-61.

72.   Carlessi, R., et al., *Mouse Models of Hepatocellular Carcinoma*, in *Hepatocellular Carcinoma*, J.E.E. Tirnitz-Parker, Editor. 2019, Codon Publications

Copyright: The Authors.: Brisbane (AU).

73.   Liu, F., Y. Song, and D. Liu, *Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA.* Gene Ther, 1999. **6**(7): p. 1258-66.

74. Zhang, G., V. Budker, and J.A. Wolff, *High levels of foreign gene expression in hepatocytes after tail vein injections of naked plasmid DNA.* Hum Gene Ther, 1999. **10**(10): p. 1735-7.

75. Suda, T. and D. Liu, *Hydrodynamic gene delivery: its principles and applications.* Mol Ther, 2007. **15**(12): p. 2063-9.

76. Chen, X. and D.F. Calvisi, *Hydrodynamic transfection for generation of novel mouse models for liver cancer research.* Am J Pathol, 2014. **184**(4): p. 912-923.

77. Ivics, Z., et al., *Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.* Cell, 1997. **91**(4): p. 501-10.

78. Yant, S.R., et al., *Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system.* Nat Genet, 2000. **25**(1): p. 35-41.

79. Ju, H.-L., et al., *Transgenic mouse models generated by hydrodynamic transfection for genetic studies of liver cancer and preclinical testing of anti-cancer therapy.* International Journal of Cancer, 2016. **138**(7): p. 1601-1608.

80. Dang, C.V., *MYC on the path to cancer.* Cell, 2012. **149**(1): p. 22-35.

81. Thorgeirsson, S.S. and E. Santoni-Rugiu, *Transgenic mouse models in carcinogenesis: interaction of c-myc with transforming growth factor alpha and hepatocyte growth factor in hepatocarcinogenesis.* Br J Clin Pharmacol, 1996. **42**(1): p. 43-52.

82. Conner, E.A., et al., *E2F1 blocks and c-Myc accelerates hepatic ploidy in transgenic mouse models.* Biochemical and Biophysical Research Communications, 2003. **302**(1): p. 114-120.

83. Calvisi, D.F., et al., *Increased lipogenesis, induced by AKT-mTORC1-RPS6 signaling, promotes development of human hepatocellular carcinoma.* Gastroenterology, 2011. **140**(3): p. 1071-83.

84. Wuestefeld, T., et al., *A Direct in vivo RNAi screen identifies MKK4 as a key regulator of liver regeneration.* Cell, 2013. **153**(2): p. 389-401.

85. Rudalska, R., et al., *In vivo RNAi screening identifies a mechanism of sorafenib resistance in liver cancer.* Nat Med, 2014. **20**(10): p. 1138-46.

86. Hedstrom, L., *Serine Protease Mechanism and Specificity.* Chemical Reviews, 2002. **102**(12): p. 4501-4524.

87. Rawlings, N.D., et al., *The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database.* Nucleic Acids Research, 2017. **46**(D1): p. D624-D632.

88. Di Cera, E., *Serine proteases.* IUBMB life, 2009. **61**(5): p. 510-515.

89. Ménard, R. and A.C. Storer, *Oxyanion hole interactions in serine and cysteine proteases.* Biol Chem Hoppe Seyler, 1992. **373**(7): p. 393-400.

90. Ahern, K. and I. Rajagopal, *BIOCHEMISTRY FREE & EASY.* 2015, DaVinci Press; 3rd Edition.

91. Affara, N.I., P. Andreu, and L.M. Coussens, *Delineating protease functions during cancer development.* Methods Mol Biol, 2009. **539**: p. 1-32.

92.   Tanabe, L.M. and K. List, *The role of type II transmembrane serine protease-mediated signaling in cancer.* Febs j, 2017. **284**(10): p. 1421-1436.

93.   Li, S., et al., *Hepsin: a multifunctional transmembrane serine protease in pathobiology.* The FEBS Journal, 2021. **288**(18): p. 5252-5264.

94.   Partanen, J.I., et al., *Tumor suppressor function of Liver kinase B1 (Lkb1) is linked to regulation of epithelial integrity.* Proc Natl Acad Sci U S A, 2012. **109**(7): p. E388-97.

95.   Wang, C.H., et al., *TMPRSS4 facilitates epithelial-mesenchymal transition of hepatocellular carcinoma and is a predictive marker for poor prognosis of patients after curative resection.* Sci Rep, 2015. **5**: p. 12366.

96.   Zhu, F., et al., *HtrA1 regulates epithelial-mesenchymal transition in hepatocellular carcinoma.* Biochem Biophys Res Commun, 2015. **467**(3): p. 589-94.

97.   Nouri, K., Y. Feng, and A.D. Schimmer, *Mitochondrial ClpP serine protease-biological function and emerging target for cancer therapy.* Cell Death Dis, 2020. **11**(10): p. 841.

98.   Sen, B., et al., *Senescent Hepatocytes in Decompensated Liver Show Reduced UPR(MT) and Its Key Player, CLPP, Attenuates Senescence In Vitro.* Cell Mol Gastroenterol Hepatol, 2019. **8**(1): p. 73-94.

99.   Gibellini, L., et al., *The biology of Lonp1: More than a mitochondrial protease.* Int Rev Cell Mol Biol, 2020. **354**: p. 1-61.

100.   *Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information.* 2004   cited 2022 Jan 24]; Available from: https://www.ncbi.nlm.nih.gov/gene/.

101.   Consortium, T.U., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Research, 2020. **49**(D1): p. D480-D489.

102.   Yates, A., et al., *Ensembl 2016.* Nucleic Acids Research, 2015. **44**(D1): p. D710-D716.

103.   Miyakoshi, K., et al., *The identification of novel ovarian proteases through the use of genomic and bioinformatic methodologies.* Biol Reprod, 2006. **75**(6): p. 823-35.

104.   Wahlberg, P., et al., *Expression and localization of the serine proteases high-temperature requirement factor A1, serine protease 23, and serine protease 35 in the mouse ovary.* Endocrinology, 2008. **149**(10): p. 5070-7.

105.   Golz, S.E., DE), Summer, Holger (Wuppertal, DE), Geerts, Andreas (Wuppertal, DE), Brüggemeier, Ulf (Leichlingen, DE), Albrecht, Barbara (Wulfrath, DE), Klein, Martina (Dusseldorf, DE), Steppan, Sonja (Neu-Isenburg, DE), Ellinghaus, Peter (Melle, DE), D'urso, Donatella (Duesseldorf, DE), Seewald, Michael (Langenfeld, DE), Milting, Hendrik (Bad Oeynhausen, DE),, *PRSS23 AS A BIOMARKER, THERAPEUTIC AND DIAGNOSTIC TARGET.* 2010: United States.

106.   Chen, I.H., et al., *PRSS23 is essential for the Snail-dependent endothelial-to-mesenchymal transition during valvulogenesis in zebrafish.* Cardiovasc Res, 2013. **97**(3): p. 443-53.

107.   LeBleu, V.S., et al., *Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis.* Nat Med, 2013. **19**(2): p. 227-31.

108.   Qiu, X., et al., *Single-cell RNA sequencing of human femoral head in vivo.* Aging (Albany NY), 2021. **13**(11): p. 15595-15619.

109. Zajchowski, D.A., et al., *Identification of Gene Expression Profiles That Predict the Aggressive Behavior of Breast Cancer Cells.* Cancer Research, 2001. **61**(13): p. 5168-5178.

110. Ifon, E.T., et al., *U94 alters FN1 and ANGPTL4 gene expression and inhibits tumorigenesis of prostate cancer cell line PC3.* Cancer Cell International, 2005. **5**(1): p. 19.

111. Jones, S., et al., *Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses.* Science, 2008. **321**(5897): p. 1801-1806.

112. Tanaka, T., et al., *Molecular Pathogenesis of Pancreatic Ductal Adenocarcinoma: Impact of miR-30c-5p and miR-30c-2-3p Regulation on Oncogenic Genes.* Cancers (Basel), 2020. **12**(10).

113. Jarząb, B., et al., *Gene Expression Profile of Papillary Thyroid Cancer: Sources of Variability and Diagnostic Implications.* Cancer Research, 2005. **65**(4): p. 1587-1597.

114. Han, B., et al., *PRSS23 knockdown inhibits gastric tumorigenesis through EIF2 signaling.* Pharmacol Res, 2019. **142**: p. 50-57.

115. Chan, H.S., et al., *Serine protease PRSS23 is upregulated by estrogen receptor α and associated with proliferation of breast cancer cells.* PLoS One, 2012. **7**(1): p. e30397.

116. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Research, 2010. **38**(16): p. e164-e164.

117. Robinson, J.T., et al., *Integrative genomics viewer.* Nature Biotechnology, 2011. **29**(1): p. 24-26.

118. Novoradovsky, A., et al., *Computational Principles of Primer Design for Site Directed Mutagenesis.* TechConnect Briefs, 2005. **1, Technical Proceedings of the 2005 NSTI Nanotechnology Conference and Trade Show**: p. 532 - 535.

119. Hayer, A., et al., *Engulfed cadherin fingers are polarized junctional structures between collectively migrating endothelial cells.* Nat Cell Biol, 2016. **18**(12): p. 1311-1323.

120. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nature Methods, 2012. **9**(7): p. 676-682.

121. *"Four Parameter Logistic Curve" online data analysis tool, MyAssays Ltd.* 07 th January 2021]; Available from: http://www.myassays.com/four-parameter-logistic-curve.assay.

122. Dai, M., et al. *Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data.* Nucleic acids research, 2005. **33**, e175 DOI: 10.1093/nar/gni179.

123. Geistlinger, L., G. Csaba, and R. Zimmer, *Bioconductor's EnrichmentBrowser: seamless navigation through combined results of set- & network-based enrichment analysis.* BMC Bioinformatics, 2016. **17**: p. 45.

124. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes.* Nat Protoc, 2006. **1**(6): p. 2856-60.

125. Tyanova, S., T. Temu, and J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nature Protocols, 2016. **11**(12): p. 2301-2319.

126. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.

127. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-342.

128. Huber, W., et al., *Variance stabilization applied to microarray data calibration and to the quantification of differential expression.* Bioinformatics, 2002. **18**(suppl_1): p. S96-S104.

129. *Replace missing values from normal distribution.* [cited 2021; Available from: http://www.coxdocs.org/doku.php?id=perseus:user:activities:matrixprocessing:imputation:replacemissingfromgaussian.

130. Stekhoven, D.J. and P. Bühlmann, *MissForest--non-parametric missing value imputation for mixed-type data.* Bioinformatics, 2012. **28**(1): p. 112-8.

131. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.

132. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.

133. Kelley, L.A. and M.J.E. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nature Protocols, 2009. **4**(3): p. 363-371.

134. Mookerjee, S.A., et al., *The contributions of respiration and glycolysis to extracellular acid production.* Biochim Biophys Acta, 2015. **1847**(2): p. 171-181.

135. Szklarczyk, D., et al., *The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets.* Nucleic Acids Res, 2021. **49**(D1): p. D605-d612.

136. Ghadially, F.N., *Ultrastructural pathology of the cell and matrix.* 4th ed. Vol. 1. 1997: Butterworth-Heinemann.

137. Anghel, S.I., et al., *Breast cancer suppressor candidate-1 (BCSC-1) is a melanoma tumor suppressor that down regulates MITF.* Pigment Cell & Melanoma Research, 2012. **25**(4): p. 482-487.

138. Zhou, Y.Q., et al., *Tumor suppressor function of BCSC-1 in nasopharyngeal carcinoma.* Cancer Sci, 2009. **100**(10): p. 1817-22.

139. Huang, J., et al., *Up-regulation of DLK1 as an imprinted gene could contribute to human hepatocellular carcinoma.* Carcinogenesis, 2007. **28**(5): p. 1094-103.

140. Yu, F., et al., *Delta-like 1 contributes to cell growth by increasing the interferon-inducible protein 16 expression in hepatocellular carcinoma.* Liver Int, 2010. **30**(5): p. 703-14.

141. Cai, C.M., et al., *Targeting endogenous DLK1 exerts antitumor effect on hepatocellular carcinoma through initiating cell differentiation.* Oncotarget, 2016. **7**(44): p. 71466-71476.

142. Huang, C.C., et al., *Delta-like 1 homologue promotes tumorigenesis and epithelial-mesenchymal transition of ovarian high-grade serous carcinoma through activation of Notch signaling.* Oncogene, 2019. **38**(17): p. 3201-3215.

143. Pittaway, J.F.H., et al., *The role of delta-like non-canonical Notch ligand 1 (DLK1) in cancer.* Endocr Relat Cancer, 2021. **28**(12): p. R271-r287.

144. Kawakami, T., et al., *Imprinted DLK1 is a putative tumor suppressor gene and inactivated by epimutation at the region upstream of GTL2 in human renal cell carcinoma.* Hum Mol Genet, 2006. **15**(6): p. 821-30.

145. Gabay, M., Y. Li, and D.W. Felsher, *MYC activation is a hallmark of cancer initiation and maintenance.* Cold Spring Harb Perspect Med, 2014. **4**(6).

146. Kumar, S. and R. Nussinov, *Salt bridge stability in monomeric proteins.* J Mol Biol, 1999. **293**(5): p. 1241-55.

147. Pollock, A.S., *Intracellular pH of hepatocytes in primary monolayer culture.* Am J Physiol, 1984. **246**(5 Pt 2): p. F738-44.

148. Berezhkovskiy, L.M., S. Wong, and J.S. Halladay, *On the maintenance of hepatocyte intracellular pH 7.0 in the in-vitro metabolic stability assay.* J Pharmacokinet Pharmacodyn, 2013. **40**(6): p. 683-9.

149. Casey, J.R., S. Grinstein, and J. Orlowski, *Sensors and regulators of intracellular pH.* Nature Reviews Molecular Cell Biology, 2010. **11**(1): p. 50-61.

150. Kokkonen, N., et al., *Abnormal Golgi pH Homeostasis in Cancer Cells Impairs Apical Targeting of Carcinoembryonic Antigen by Inhibiting Its Glycosyl-Phosphatidylinositol Anchor-Mediated Association with Lipid Rafts.* Antioxid Redox Signal, 2019. **30**(1): p. 5-21.

151. Zhuang, Y. and W.K. Miskimins, *Cell cycle arrest in Metformin treated breast cancer cells involves activation of AMPK, downregulation of cyclin D1, and requires p27Kip1 or p21Cip1.* J Mol Signal, 2008. **3**: p. 18.

152. Jones, R.G., et al., *AMP-activated protein kinase induces a p53-dependent metabolic checkpoint.* Mol Cell, 2005. **18**(3): p. 283-93.

153. Hannan, K.M., et al., *RNA polymerase I transcription in confluent cells: Rb downregulates rDNA transcription during confluence-induced cell cycle arrest.* Oncogene, 2000. **19**(31): p. 3487-3497.

154. Rattan, R., et al., *5-Aminoimidazole-4-carboxamide-1-beta-D-ribofuranoside inhibits cancer cell proliferation in vitro and in vivo via AMP-activated protein kinase.* J Biol Chem, 2005. **280**(47): p. 39582-93.

155. Kim, K.H. and J.M. Sederstrom, *Assaying Cell Cycle Status Using Flow Cytometry.* Curr Protoc Mol Biol, 2015. **111**: p. 28.6.1-28.6.11.

156. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

157. Winkler, J., et al., *Concepts of extracellular matrix remodelling in tumour progression and metastasis.* Nat Commun, 2020. **11**(1): p. 5120.

158. Tagliabracci, Vincent S., et al., *A Single Kinase Generates the Majority of the Secreted Phosphoproteome.* Cell, 2015. **161**(7): p. 1619-1632.

159. Griss, J., et al., *ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis.* Mol Cell Proteomics, 2020. **19**(12): p. 2115-2125.

160. Gonzalez-Begne, M., et al., *Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT).* J Proteome Res, 2009. **8**(3): p. 1304-14.

161.   Robinson, G.L., et al., *Switching from aerobic glycolysis to oxidative phosphorylation modulates the sensitivity of mantle cell lymphoma cells to TRAIL.* Oncogene, 2012. **31**(48): p. 4996-5006.

162.   Hardie, D.G., et al., *Management of cellular energy by the AMP-activated protein kinase system.* FEBS Letters, 2003. **546**(1): p. 113-120.

163.   Dai, M.S. and H. Lu, *Crosstalk between c-Myc and ribosome in ribosomal biogenesis and cancer.* J Cell Biochem, 2008. **105**(3): p. 670-7.

164.   Annunziata, I., et al., *MYC competes with MiT/TFE in regulating lysosomal biogenesis and autophagy through an epigenetic rheostat.* Nature Communications, 2019. **10**(1): p. 3623.

165.   La Spina, M., et al., *MiT/TFE Family of Transcription Factors: An Evolutionary Perspective.* Frontiers in Cell and Developmental Biology, 2021. **8**(1580).